PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/10, C07K 14/705 — A2

(11) International Publication Number: WO 00/52153

(43) International Publication Date: 8 September 2000 (08.09.00)

(21) International Application Number: PCT/US00/05764

(22) International Filing Date: 1 March 2000 (01.03.00)

(30) Priority Data:
60/122,569 2 March 1999 (02.03.99) US

(71) Applicant *(for all designated States except US)*: MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HOWARD, Russell, J. [AU/US]; 12700 Viscaino Road, Los Altos Hills, CA (US). PATTEN, Phillip, A. [US/US]; 2680 Fayette Drive, Apt. 506, Mountain View, CA 94040 (US).

(74) Agents: SMITH, Timothy, L. et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SURROGATE ORPHAN LIGANDS FOR ORPHAN RECEPTORS

(57) Abstract

This invention provides methods for obtaining surrogate ligands for orphan receptors, as well as surrogate receptors for orphan ligands. Themethods are also useful for obtaining optimized ligands and/or receptors that exhibit an enhanced ability to modulate a biological activity compared to a naturally occurring cognate receptor or cognate ligand.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SURROGATE ORPHAN LIGANDS FOR ORPHAN RECEPTORS

BACKGROUND OF THE INVENTION

Field of the Invention

This invention pertains to the field of obtaining surrogate ligands that are functional upon orphan receptors.

Background

Rapid genomic DNA sequencing often uncovers new receptors, termed "orphan receptors," for which the cognate natural ligand(s) are unknown. Newly discovered orphan receptors are often assignable to a family of existing receptors for which one or more ligands may have already been identified and cloned, and the receptor-ligand interactions studied. Existing members of the ligand family, however, often show little or no binding or biological activity towards a new putative member of the receptor family. The elucidation of the biological function of an orphan receptor must generally await the identification and characterization of the natural cognate ligand for the orphan receptor. Similarly, upon the discovery of a previously unknown ligand, the elucidation of its biological function must await identification of its cognate receptor.

Previously available approaches for identifying cognate ligands for an orphan receptor, and cognate receptors for an orphan ligand, suffer from serious drawbacks. The approach of rapidly cloning as many possible members of a ligand or receptor family by homology, for example, is likely to prove slow and tortuous due to the often large number of ligands or receptors in a ligand or receptor family. Thus, a need exists for improved methods by which one can obtain a ligand that binds to and exerts a biological activity upon an orphan receptor. The present invention fulfills this and other needs.

SUMMARY OF THE INVENTION

In a first embodiment, the present invention provides methods for obtaining a surrogate ligand for an orphan receptor. The methods involve: (1) creating a library of

recombinant polynucleotides; and (2) screening the library to identify a recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor and/or modulate the activity of the orphan receptor.

In presently preferred embodiments, a library of recombinant polypeptides is obtained by recombining at least first and second forms of a nucleic acid, each of which forms encodes a ligand for a member of a receptor family, or a fragment of said ligand, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant nucleic acids. The receptor family is chosen based upon homology to the orphan receptor of interest. The library of recombinant nucleic acids is then screened to identify a recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor and/or modulate the activity of the orphan receptor.

In some embodiments, these methods further involve: (3) recombining at least one recombinant polynucleotide that encodes a surrogate ligand identified in the first round of screening with a further form of the nucleic acid, which is the same or different from the first and second forms, to produce a further library of recombinant polynucleotides; and (4) screening the further library to identify at least one further optimized recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor and/or modulate the activity of the receptor. The recombining and screening steps are repeated, as necessary, until the surrogate ligand encoded by the further optimized recombinant polynucleotide exhibits an enhanced ability to specifically bind to the ligand binding domain of the orphan receptor.

In other embodiments, the screening methods involve expressing the library of recombinant polynucleotides, and contacting the resulting library of candidate surrogate ligands with a test cell that contains a polypeptide which comprises: a) a ligand binding domain of the orphan receptor (which can be an extracellular domain of the receptor); and b) a cytoplasmic and/or DNA binding domain of a second receptor, whereby the binding of a ligand to the ligand binding domain of the peptide results in a detectable effect on the test cells. The surrogate ligand typically exhibits an agonist function upon binding to the ligand binding domain of the orphan receptor, although in some cases an antagonist effect is

observed. The second receptor is, in some embodiments, a cytokine receptor such as, for example, an interleukin receptor, an interferon receptor, a chemokine receptor, a hematopoietic growth factor receptor, a tumor necrosis factor receptor, and a transforming growth factor. The DNA binding domain can also be obtained from the orphan ligand itself (*i.e.*, the entire orphan ligand is used in the screening assay).

The invention also provides methods of identifying a surrogate ligand by expressing a library of recombinant polynucleotides to obtain a library of candidate surrogate ligands, and screening the candidate surrogate ligands using a reporter gene system. For example, the candidate surrogate ligands can be contacted with a test cell that includes:

a) a fusion polypeptide comprising: 1) a ligand binding domain of the orphan receptor; and 2) a DNA binding domain of a second receptor; and

b) a reporter gene construct which comprises a response element to which the DNA binding domain can bind, wherein the response element is operably linked to a promoter that is operative in the cell and the promoter is operably linked to a reporter gene. The screening involves determining whether the reporter gene is expressed at a higher or lower level in the presence of a candidate surrogate ligand compared to expression in the absence of the candidate surrogate ligand. In these embodiments, the DNA binding domain can be, for example, a GAL4 DNA binding domain, or can be obtained from a receptor such as, for example, an estrogen receptor, a progesterone receptor, a glucocorticoid receptor, an androgen receptor, a mineralcorticoid receptor, a vitamin D receptor, a retinoid receptor, and a thyroid hormone receptor, or can be from the orphan receptor itself if a response element for the orphan receptor is known.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A and 1B show the amino acid sequences and genealogies of shuffled human interferons. Figure 1A shows the amino acid sequences of seven evolved IFN-αs and the eight native Hu-IFN-αs from which they are derived are shown. The most parsimonious genealogies of the shuffled IFN-αs are shown schematically. Recombination junctions are shown at the midpoint between two amino acids derived from different parental genes. The gene segments are colored according to which parental gene they are derived from (Hu-IFN-α1, red; Hu-IFN-α5, green; Hu-IFN-α8, yellow; Hu-IFN-α16, purple; Hu-

IFN-α17, orange; Hu-IFN-αF, blue; Hu-IFN-αH, gray). Amino acids that arose by point mutation during DNA shuffling are circled.

Figure 1B shows the amino acid sequence of one of the cycle two chimeras, IFN-α-CH2.2, which is aligned with the most potent human and mouse IFN-αs, Hu-IFN-α1 and Mu-IFN-α4. The IFN-α residues that putatively contact the IFN-α receptor (Fish, E. N. (1992) *J. Interferon Res.* **12(4)**:257-66; Uze *et al.* (1994) *J. Mol. Biol.* **243(2)**: 245-57) are boxed. Residues in Hu-IFN-α1 that have been shown by site directed mutagenesis to contribute to activity on mouse cells (Horisberger, M. A., and Di Marco, S. (1995) *Pharmacol. Ther.* **66(3)**: 507-3411; Weber *et al.* (1987) *EMBO. J.* **6(3)**:591-8; Fish, *supra.*, Uze *et al.*, *supra.*) are shaded.

Figure 2 shows the antiviral activities of native IFN-αs and an evolved IFN-α. The results from the antiviral assay on murine L929 cells of Hu-IFN-α2a, Hu-IFN-α1, Mu-IFN- α4 and IFN-α-CH2.1 are shown. The dashed lines indicate the IFN-α dose corresponding to half-maximal protection (one unit/ml). The assays were done in triplicate and the standard errors (% of the estimated Units; Table 1) are: Mu-IFN-α4, 24%; Hu-IFN-α1, 6%; Hu-IFN-α2a, 17%; IFN-α-CH2.1, 15%.

Figure 3 shows a summary of the antiviral activities of native and evolved IFN-αs on murine L929 cells. The antiviral activities of purified CHO protein for native Mu-IFN-αs, native Hu-IFN-αs and evolved IFN-αs on murine L929 cells are shown. One unit of activity corresponds to half-maximal protection from a lethal ECMV viral challenge. The arrows on the right indicate the fold improvement of IFN-α-CH2.3 relative to Hu-IFN-α1 and Hu-IFN-α2a. The activities of the proteins were measured in four independent experiments, and the rank orders of the clones is the same in all four assays, with the exception of assay #3 in which Mu-IFN-α4 exceeded the activity of IFN-α-CH1.1, but not the round two evolved IFN-αs.

Figure 4 provides a structural modeling model of the alpha carbon backbone of IFN-α-CH2.2, based on the NMR structure of Hu-IFN-α2a (Scarozza *et al.* (1992) *J. Interferon Res.* **12**: 35-42). The protein backbone is colored to indicate the native Hu-IFN-α segment from which it is derived (Residues 29-39, 121-140 Hu-IFN-α1, red; Residues 46-120 Hu-IFN-α5, green; Residues 40-45 Hu-IFN-α8, yellow; Residues 1-28 Hu-IFN-αF,

blue; Residues 141-166 Hu-IFN-αH, gray). The side chains of putative murine IFN-α receptor contacting residues K121 and R125 are shown.

DETAILED DESCRIPTION

Definitions

The term "cytokine" includes, for example, interleukins, interferons, chemokines, hematopoietic growth factors, tumor necrosis factors and transforming growth factors. In general these are small molecular weight proteins that regulate maturation, activation, proliferation and differentiation of the cells of the immune system.

A "surrogate ligand" is a polypeptide that can bind to a receptor for which the surrogate ligand is not a naturally occurring cognate ligand, and thus typically mediate a biological effect. In some instances, the receptor to which the surrogate ligand binds is an orphan receptor for which no cognate ligand is known; in other instances, the receptor has one or more known cognate ligands but the surrogate receptor has a differential binding and/or biological mediating effect compared to a naturally occurring cognate ligand. Conversely, a "surrogate receptor" is a polypeptide that can act as a receptor for a ligand for which the polypeptide is not a naturally occurring cognate receptor. Again, the ligand can be an orphan ligand for which no known cognate receptors are known, or can be a ligand for which one or more cognate receptors are known but which exhibits a differential binding and/or biological mediating effect compared to a naturally occurring cognate receptor.

An "orphan receptor" is a putative receptor polypeptide for which a naturally occurring cognate ligand is not known at the time of the development of a surrogate ligand. Similarly, an "orphan ligand" is a putative ligand polypeptide that is believed to exhibit binding affinity for a receptor, and thus mediation of a biological effect, where the receptor is not known at the time a surrogate receptor is obtained using the methods of the invention.

An orphan receptor or an orphan ligand is said to exhibit homology to a known receptor or ligand, respectively, when the orphan receptor or ligand has one or more features that distinguish the known receptor or ligand from receptors or ligands of other families. For example, the orphan receptor can have a high degree of amino acid sequence similarity to the known over all or part of the polypeptide. Generally, when an orphan

receptor is classified on the basis of amino acid sequence similarity, the orphan receptor will be at least about 60% identical to the amino acid sequence of a corresponding domain of at least one member of a known receptor family. More preferably, the orphan receptor will be at least about 70% identical, still more preferably at least about 80% identical, and even more preferably at least about 90% identical to the corresponding domain of the known receptor. Another way to identify whether an orphan receptor exhibits homology to a known receptor (or an orphan ligand exhibits homology to a known ligand) is by determining whether the orphan receptor or ligand shares a primary sequence motif with members of a family of known receptors or ligands. Motifs of different receptor families are well known to those of skill in the art (*e.g.*, C-X-C, C-C for chemokines). Yet another indication that an orphan receptor might belong to a particular receptor family is that the structure of the orphan receptor shares features with the known receptors. For example, an Ig fold, an MHC fold, and the like, can provide information as to which family of receptors an orphan receptor is likely to be a member.

The term "screening" describes, in general, a process that identifies polypeptides that function as surrogate ligands or surrogate receptors. Several properties of the respective molecules can be used in selection and screening including, for example, ability to bind to a ligand binding domain of the orphan receptor. The binding is preferably accompanied by modulation of an activity (*e.g.*, enhanced or reduced expression of a reporter gene that is responsive to a DNA binding domain or intracellular domain of a second receptor to which the orphan receptor ligand binding domain is attached. Selection is a form of screening in which identification and physical separation are achieved simultaneously by expression of a selection marker, which, in some genetic circumstances, allows cells expressing the marker to survive while other cells die (or vice versa). Screening markers include, for example, luciferase, beta-galactosidase and green fluorescent protein. Selection markers include drug and toxin resistance genes, and the like. Although spontaneous selection can and does occur in the course of natural evolution, in the present methods selection is performed by man.

A "exogenous DNA segment", "heterologous sequence" or a "heterologous nucleic acid", as used herein, is one that originates from a source foreign to a particular host

cell, or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell, but has been modified. Modification of a heterologous nucleic acid in the applications described herein typically occurs through the use of DNA shuffling. Thus, the terms refer to a DNA segment which is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell genome at which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides (*i.e.*, polypeptides that are not native to the host cell, or are native to the host cell but are in modified form compared to the natural form of the polypeptide).

The term "isolated", when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. In particular, an "isolated gene" or an "isolated nucleic acid" is separated from open reading frames which flank the gene in its natural chromosomal location and encode a protein other than the gene of interest. An "isolated" polypeptide or nucleic acid is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein or nucleic acid which is the predominant species present in a preparation is said to be "substantially purified." The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least about 50% pure, more preferably at least about 85% pure, and most preferably at least about 99% pure.

The term "naturally-occurring" is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses, bacteria, protozoa, insects, plants or mammalian tissue) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides which

have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.* degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.* (1991) *Nucleic Acid Res. 19*: 5081; Ohtsuka *et al.* (1985) *J. Biol. Chem. 260:* 2605-2608; Cassol *et al.* (1992) ; Rossolini *et al.* (1994) *Mol. Cell. Probes 8*: 91-98).

The term nucleic acid is used interchangeably with gene, cDNA, and mRNA. Accordingly, the term "gene" is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

"Nucleic acid derived from a gene" refers to a nucleic acid for whose synthesis the gene, or a subsequence thereof, has ultimately served as a template. Thus, an mRNA, a cDNA reverse transcribed from an mRNA, an RNA transcribed from a gene or cDNA, a DNA amplified from the gene or cDNA, an RNA transcribed from the amplified DNA, *etc.*, are all derived from the gene and detection of such derived products is indicative of the presence and/or abundance of the original gene and/or gene transcript in a sample.

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it increases the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several

kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous.

The term "recombinant" when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, or expresses a peptide or protein encoded by a heterologous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also contain genes found in the native form of the cell wherein the genes are modified and re-introduced into the cell by artificial means. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques.

A "recombinant expression cassette" or simply an "expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with nucleic acid elements that are capable of effecting expression of a structural gene in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes a nucleic acid to be transcribed (*e.g.*, a member of a library of recombinant polynucleotides), and a promoter. Additional factors necessary or helpful in effecting expression may also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell. Transcription termination signals, enhancers, and other nucleic acid sequences that influence gene expression, can also be included in an expression cassette.

A "recombinant polynucleotide" or a "recombinant polypeptide" is a non-naturally occurring polynucleotide or polypeptide that includes nucleic acid or amino acid sequences, respectively, from more than one source nucleic acid or polypeptide, which source nucleic acid or polypeptide can be a naturally occurring nucleic acid or polypeptide, or can itself have been subjected to mutagenesis or other type of modification. The source polynucleotides or polypeptides from which the different nucleic acid or amino acid sequences are derived are sometimes homologous (*i.e.*, have, or encode a polypeptide that encodes, the same or a similar structure and/or function), and are often from different

isolates, serotypes, strains, species, of organism or from different disease states, for example. A recombinant ligand, for example, will have amino acids from more than one naturally occurring ligand.

The terms "identical" or percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to two or more sequences or subsequences that have at least 60%, preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In some embodiments, the sequences are substantially identical over a particular domain (*e.g.*, an extracellular or intracellular domain, or a DNA binding domain or ligand binding domain), or are substantially identical over the entire length of the coding regions.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA,

and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally* Ausubel *et al., infra*).

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al., J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an

indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to", refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target polynucleotide sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Typically, under "stringent conditions" a probe will hybridize to its target subsequence, but to no other sequences.

The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50%

formamide with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see,* Sambrook, *infra.*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 1x SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 4-6x SSC at 40°C for 15 minutes. For short probes (*e.g.*, about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.0 M Na^+ ion, typically about 0.01 to 1.0 M Na^+ ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

A further indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with, or specifically binds to, the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions.

A "specific binding affinity" between two molecules, for example, a ligand and a receptor, means a preferential binding of one molecule for another in a mixture of molecules. The binding of the molecules can be considered specific if the binding affinity is about 1×10^4 M^{-1} to about 1×10^6 M^{-1} or greater.

The phrase "specifically (or selectively) binds to" or "specifically (or selectively) immunoreactive with", when referring to a protein or peptide (*e.g.*, a ligand), refers to a binding reaction which is determinative of the presence of the protein, or an

epitope from the protein, in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated assay conditions, the specified ligands bind to a particular receptor (*e.g.*, an orphan receptor or an antibody) and do not bind in a significant amount to other proteins present in the sample. Antibodies raised against a multivalent antigenic polypeptide will generally bind to the proteins from which one or more of the epitopes were obtained. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. *See* Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York "Harlow and Lane"), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

"Conservatively modified variations" of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every polynucleotide sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. *See, e.g.,* Creighton (1984) *Proteins*, W.H. Freeman and Company, for additional groupings of amino acids. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations".

A "subsequence" refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (*e.g.,* polypeptide) respectively.

Description of the Preferred Embodiments

The present invention provides methods for obtaining ligands for receptors, in particular receptors for which cognate ligands are not yet known. The methods are also useful for obtaining recombinant ligands that exhibit greater or reduced binding affinity for, and/or biological activation of, a known receptor, compared to the naturally occurring cognate ligand for the receptor. Conversely, the methods are also useful for obtaining a receptor for a ligand for which a cognate receptor is not yet known, or for which a receptor that has greater or reduced binding affinity for, and/or biological activation of, a known ligand.

The methods of the invention provide significant advantages over previously available methods of identifying ligands for newly discovered receptors, or receptors for newly discovered ligands. Unlike previously available methods, the surrogate ligands or surrogate receptors can be obtained relatively quickly, using a relatively small number of assays. The methods are scalable and generic, so they can rapidly and economically be applied to any receptor family of interest to obtain variants that have novel properties. Moreover, little or no structural information regarding the interaction between ligand and receptor is necessary in order to obtain the surrogate ligands.

The methods of the invention for obtaining a surrogate ligand for an orphan receptor involve creating a library of recombinant polynucleotides, which library is then screened to identify a recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor. The creation of recombinant libraries, as well as screening methods are described below.

A. Creation of Recombinant Libraries

The invention involves creating recombinant libraries of polynucleotides that are then screened to identify those library members that exhibit a desired property, *e.g.*, ability to act as a surrogate ligand for an orphan receptor, or as a surrogate receptor for an orphan ligand. The recombinant libraries can be created using any of various methods, as described below.

Methods for obtaining recombinant polynucleotides and/or for obtaining diversity in nucleic acids used as the substrates for DNA shuffling as described herein include, for example, homologous recombination (PCT/US98/05223; Publ. No. WO98/42727); oligonucleotide-directed mutagenesis (for review *see*, Smith, *Ann. Rev. Genet.* 19: 423-462 (1985); Botstein and Shortle, *Science* 229: 1193-1201 (1985); Carter, *Biochem. J.* 237: 1-7 (1986); Kunkel, "The efficiency of oligonucleotide directed mutagenesis" *in Nucleic acids & Molecular Biology*, Eckstein and Lilley, eds., Springer Verlag, Berlin (1987)). Included among these methods are oligonucleotide-directed mutagenesis (Zoller and Smith, *Nucl. Acids Res.* 10: 6487-6500 (1982), *Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 329-350 (1987)) phosphothioate-modified DNA mutagenesis (Taylor *et al.*, *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor *et al.*, *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye and Eckstein, *Nucl. Acids Res.* 14: 9679-9698 (1986); Sayers *et al.*, *Nucl. Acids Res.* 16: 791-802 (1988); Sayers *et al.*, *Nucl. Acids Res.* 16: 803-814 (1988)), mutagenesis using uracil-containing templates (Kunkel, *Proc. Nat'l. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel *et al.*, *Methods in Enzymol.* 154: 367-382)); mutagenesis using gapped duplex DNA (Kramer *et al.*, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer and Fritz, *Methods in Enzymol.* 154: 350-367 (1987); Kramer *et al.*, *Nucl. Acids Res.* 16: 7207 (1988)); and Fritz *et al.*, *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional suitable methods include point mismatch repair (Kramer *et al.*, *Cell* 38:

879-887 (1984)), mutagenesis using repair-deficient host strains (Carter *et al.*, *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1987)), deletion mutagenesis (Eghtedarzadeh and Henikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction-selection and restriction-purification (Wells *et al.*, *Phil. Trans. R. Soc. Lond.* A 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar *et al.*, *Science* 223: 1299-1301 (1984); Sakamar and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells *et al.*, *Gene* 34: 315-323 (1985); and Grundström *et al.*, *Nucl. Acids Res.* 13: 3305-3316 (1985). Kits for mutagenesis are commercially available (*e.g.*, Bio-Rad, Amersham International, Anglian Biotechnology).

In a presently preferred embodiment, the recombinant libraries are prepared using DNA shuffling. The shuffling and screening or selection can be used to "evolve" individual genes, whole plasmids or viruses, multigene clusters, or even whole genomes (Stemmer (1995) *Bio/Technology* 13:549-553). Reiterative cycles of recombination and screening/selection can be performed to further evolve the nucleic acids of interest. Such techniques do not require the extensive analysis and computation required by conventional methods for polypeptide engineering. Shuffling allows the recombination of large numbers of mutations in a minimum number of selection cycles, in contrast to traditional, pairwise recombination events. Thus, the sequence recombination techniques described herein provide particular advantages in that they provide recombination between mutations in any or all of these, thereby providing a very fast way of exploring the manner in which different combinations of mutations can affect a desired result. In some instances, however, structural and/or functional information is available which, although not required for sequence recombination, provides opportunities for modification of the technique.

Exemplary formats and examples for sequence recombination, sometimes referred to as DNA shuffling, evolution, or molecular breeding, have been described by the present inventors and co-workers in co-pending applications U.S. Patent Application Serial No. 08/198,431, filed February 17, 1994, Serial No. PCT/US95/02126, filed, February 17, 1995, Serial No. 08/425,684, filed April 18, 1995, Serial No. 08/537,874, filed October 30, 1995, Serial No. 08/564,955, filed November 30, 1995, Serial No. 08/621,859, filed March 25, 1996, Serial No. 08/621,430, filed March 25, 1996, Serial No. PCT/US96/05480, filed

April 18, 1996, Serial No. 08/650,400, filed May 20, 1996, Serial No. 08/675,502, filed July 3, 1996, Serial No. 08/721, 824, filed September 27, 1996, Serial No. PCT/US97/17300, filed September 26, 1997, and Serial No. PCT/US97/24239, filed December 17, 1997; Stemmer, *Science* 270:1510 (1995); Stemmer *et al.*, *Gene* 164:49-53 (1995); Stemmer, *Bio/Technology* 13:549-553 (1995); Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91:10747-10751 (1994); Stemmer, *Nature* 370:389-391 (1994); Crameri *et al.*, *Nature Medicine* 2(1):1-3 (1996); Crameri *et al.*, *Nature Biotechnology* 14:315-319 (1996), each of which is incorporated by reference in its entirety for all purposes.

The methods require at least two variant forms of a starting substrate, such as a nucleic acid that encodes a receptor, or a part of a receptor if a surrogate ligand is desired. The variant forms of candidate substrates can show substantial sequence or secondary structural similarity with each other, but they should also differ in at least two positions. The initial diversity between forms can be the result of natural variation, *e.g.*, the different variant forms (homologs) are obtained from different individuals or strains of an organism (including geographic variants) or constitute related sequences from the same organism (*e.g.*, allelic variations). Alternatively, the initial diversity can be induced, *e.g.*, the second variant form can be generated by error-prone transcription, such as an error-prone PCR or use of a polymerase which lacks proof-reading activity (*see* Liao (1990) *Gene* 88:107-111), of the first variant form, or, by replication of the first form in a mutator strain. The initial diversity between substrates is greatly augmented in subsequent steps of recursive sequence recombination.

Sequence recombination can be achieved in many different formats and permutations of formats, which share some common principles. Recursive sequence recombination entails successive cycles of recombination to generate molecular diversity. That is, one creates a family of nucleic acid molecules showing some sequence identity to each other but differing in the presence of mutations. In any given cycle, recombination can occur *in vivo* or *in vitro*, intracellular or extracellular. Furthermore, diversity resulting from recombination can be augmented in any cycle by applying prior methods of mutagenesis (*e.g.*, error-prone PCR or cassette mutagenesis) to either the substrates or products for recombination. In some instances, a new or improved property or characteristic can be

achieved after only a single cycle of *in vivo* or *in vitro* recombination, as when using different, variant forms of the sequence, as homologs from different individuals or strains of an organism, or related sequences from the same organism, as allelic variations.

Often, improvements are achieved after one round of recombination and selection. However, recursive sequence recombination can be employed to achieve still further improvements in a desired property, such as binding affinity for an orphan receptor and/or modulation of receptor activity.

In a presently preferred embodiment, "family shuffling" is used to create the library of recombinant polynucleotides. In family shuffling, nucleic acids that encode homologous polypeptides from different strains, species, or gene families are used as the different forms of the nucleic acids. The nucleic acids can encode, for example, human and mouse homologs of a particular ligand (*e.g.*, the same ligand), or different human homologs of a ligand (*e.g.*, ligands for different receptors within a receptor family). Or the different forms of the nucleic acid can encode different ligands within a family, as well as homologs from different species. As genomics provides an increasing amount of sequence information, it is increasingly possible to directly amplify homologs with designed primers. For example, given the sequence of interferon-α genes from several species, one can design primers for amplification of the homologs. The resulting fragments can then be subjected to shuffling.

The substrate nucleic acids that are used to create the recombinant library of polynucleotides are chosen depending upon the particular application. For example, where a surrogate ligand is desired for an orphan receptor that is believed to be a member of a cytokine receptor family, polynucleotides that encode all or part of a cognate ligand for receptors of that cytokine receptor family are subjected to recombination. For example, where the orphan receptor appears to be a member of the cytokine/hematopoietic growth factor (Type I) cytokine receptor family, the starting polynucleotides can encode all or part of an IL-2, IL-4, or IL-6 polypeptide. Similarly, for an orphan receptor that appears to be a member of the interferon (Type II) receptor family, nucleic acids that encode one or more of interferon-α, interferon-β, or interferon-τ can be used as a starting substrate. For an orphan receptor of the TNF (Type III) receptor family, the starting substrates can be, for example, polynucleotides that encode tumor necrosis factor. Surrogate ligands for the Ig superfamily

of cytokine receptors can be obtained by using IL-1-encoding polynucleotides to make the recombinant library, while obtaining surrogate ligands for an orphan receptor of the seven transmembrane helix family can involve making a recombinant library using IL-8-encoding polynucleotides as the starting material.

The methods can also be used to obtain a surrogate ligand, or an improved ligand, for a member of a receptor family such as androgen receptors, estrogen receptors, glucocorticoid receptors, mineralcorticoid receptors, progesterone receptors, retinoic acid receptors, and thyroid hormone receptors, and the like. As discussed above, polynucleotides that encode one or more cognate ligands for receptors in the particular family of interest are used to create a library of recombinant polynucleotides, which is then screened to identify those recombinant polynucleotides that encode a ligand that has specific affinity for the orphan receptor of interest.

Representative, but not limiting, examples of gene families of interest, and representative ligands that can be shuffled to obtain surrogate ligands for orphan receptors, are listed in Table 1.

Table 1

Gene Family	Representative Ligands	Receptor Distribution	Receptor Type	Assay/Screen
Cysteine knot family	TGFβ, NGF, PDGFβ	Broad	Heterotypic complex	Proliferation, reporter cells
FGF Family	FGF, IL-1α, -1β	Broad	Complex role of heparan sulphates	Proliferation, reporter cells
Cysteine-cysteine chemokine family	MCP-1, MCP-2, MCP-3, MCP-1,3, MIP-1α, MIP-1β, RANTES, MIP-1α,β, I-309	Broad	7-transmembrane	Proliferation, reporter cells
Cysteine-X-cysteine chemokine family	GROα, GROβ, ENA-78, NAP-2, NAP-4, IL-8, PF-4	Broad	7-transmembrane	Proliferation, reporter cells
Short-chain four-helix bundle family	IL-2,3,4,5,7,9,13, GM-CSF, M-CSF, SCF, Flt-3 ligand	Blood cells	Homo- and heterotypic, sometimes shared	Proliferation, reporter cells
Long-chain Four-helix bundle family	IL-6,12, G-CSF, IPO, LIF, ONC, CNTF, GH, PRL, IL-11, GPA, IFNα,β, oncostatin M, cardiotrophin-1	Blood cells	Homo- and heterotypic, sometimes shared	Proliferation, reporter cells
TNF family	TNF-α, LT-α, LT-β	Broad	Single receptor, sometimes shared	Proliferation, reporter cells
Interferon family	IFN-α,β,τ,ο	Broad	Complex signaling single receptor	Proliferation, AV, reporter cells
Fc Domains of Antibodies		Broad	MHC structural homolog	Reporter cells

1. Chemokines

In some embodiments, the invention provides methods of obtaining surrogate ligands for orphan receptors that exhibit homology to one or more types of chemokine receptor. These methods involve identifying a known chemokine receptor that exhibits homology (*e.g.*, amino acid sequence similarity, conserved amino acid residues, structural similarity, and the like) to the orphan receptor. Nucleic acids that encode all or part of one or more known ligands for this known receptor are then subjected to DNA shuffling. For example, if the orphan receptor exhibits homology to a Cysteine- Cysteine (C- C) chemokine receptor (*e.g.*, CCR-1, -2, -3, -4, -5, -6, -7, -8; *see* Table 2 for examples of gene names), the shuffled ligand-encoding nucleic acids can be selected from those listed in Table 3. A shuffling reaction can involve two or more homologs of the same gene from different mammals (*e.g.*, human SCYA1 shuffled with mouse SCYA1), two or more different genes from a single mammalian species (*e.g.*, human SCYA1 shuffled with human SCYA2), or any combination thereof.

Table 2: C-C Chemokine Receptors

Gene Symbol		Gene Name
(Human)	(Mouse)	
CCR1	Cmkbr1	CKR1, CMKBR1, MIP-1a/RANTES-R, HM145, LD78-R
(CCR1L1)	Cmbkr111	CMKBR1L1, MIP-1a-R-like 1
CCR2	Cmkbr2	CKR2, CMKBR2, CCR2A, CCR2B, MCP1-R, JE/FIC-R
CCR3	Cmkbr112	CKR3, CMKBR3, eotaxin-R, CMBKR1L2, MIP-1a-R-like 2
CCR4	Cmkbr4	CKR4, CMKBR4, K5-5, MIPR17
CCR5	Cmkbr5	CKR5, CMKBR5, ChemR13, MIP-1a-R 2
CCR6	-	CMKBR6, STRL22, GPCR29, CKR-L3, GPR-CY4, DRY6, KY411, LARC-R
CCR7	Cmkbr7	CKR7, CMKBR7, EBI1 BLR2, MIP-3b-R
CCR8	Cmkbr8	CKR-8, CMKBR8, TER1, ChemR1, CKR-L1, GPR-CY6, CMKBRL2
CCR9	Cmkbr10	GPR-9-6
GPR2	-	GPR2, SEL226b

Table 3: C-C Chemokines

Gene Symbol (Human)	Gene Symbol (Mouse)	Gene Name
SCYA1	Scya1	CCL1, I-309, TCA3, P500, SISe
SCYA2	Scya2	CCL2, MCP-1, MCAF, JE, SMC-CF, GDCF-2
SCYA3, SCYA3L1	Scya3	CCL3, CCL3L1, LD78a, LD78b, AT464.1, AT464.2, G0S19-1, G0S19-2, MIP-1a, SCLTY-5, L2G25B, MIP-1aS, MIP-1aP, SISa, SISb
SCYA3L2	-	LD78g, G0S19-3(pseudogene)
SCYA4, SCYA4L	Scya4	CCL4, CCL4L, AT744.1, AT744.2, Act-2, G-26, HC21, H400, MIP-1b, LAG-1
SCYA5	Scya5	CCL5, RANTES, SISd
SCYA6	Scya6	C10, MRP-1
SCYA7	Scya7	CCL7, MCP-3, NC28, FIC, MARC
SCYA8	Scya8	CCL8, MCP-2, HC14
(SCYA9, SCYA10)	Scya9, Scya10	MRP-2, CCF18, MIP-1g
SCYA11	Scya11	CCL11, eotaxin
(SCYA12)	Scya12	CCL12, MCP-5
SCYA13	-	CCL13, MCP-4, NCC-1, CKb10
SCYA14	-	CCL14, HCC-1, HCC-3, NCC-2, CKb1, MCIF
SCYA15	-	CCL15, HCC-2, NCC-3, MIP-5, Lkn-1, MIP-1d
SCYA16	Scya16-ps	CCL16, NCC-4, LEC, HCC-4, LMC, LCC-1, CKb12
SCYA17	Scya17	CCL17, TARC, ABCD-2
SCYA18	-	CCL18, DC-CK1, PARC, MIP-4, AMAC-1, CKb7
SCYA19	-	CCL19, ELC, MIP-3b, exodus-3, CKb11
SCYA20	Scya20	CCL20, MIP-3a, LARC, exodus-1, ST38, CKb4
SCYA21	Scya21a, b	CCL21, SLC, 6Ckine, exodus-2, TCA4, 6Ckine-ser (Scya21a), 6Ckine-leu (Scya21b), CKb9
SCYA22	Scya22	CCL22, MDC, STCP-1, ABCD-1, DC/B-CK
SCYA23	-	CCL23, MIP-3, MPIF-1, CKb8, CKb8-1
SCYA24	-	CCL24, MPIF-2, CKb6, eotaxin-2
SCYA25	Scya25	CCL25, TECK, Ckb15
SCYA26	-	CCL26, SCYA26, eotaxin-3, IMAC
SCYA27	Scya27	CCL27, ALP, skinkine, ILC, ESkine, PESKY, CTAK
(clone 391)	-	clone 391

Gene Symbol		Gene Name
(Carp CC-1)	-	carp CC Chemokine-1

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/CFC/CK/CCG/CCG.html)

To obtain a surrogate ligand for an orphan receptor that exhibits homology to a Cysteine-X-Cysteine (C-X-C) chemokine receptor (*e.g.*, CXCR-1, -2, -3, or -4, and others listed in Table 4), the nucleic acids that are subjected to shuffling can include one or more of those listed in Table 5.

Table 4: C-X-C Chemokine Receptors

Gene Symbol		Gene Name
(Human)	(Mouse)	
IL8RA	-	CXCR1, CMKAR1, IL8RA, IL8R1, CDW128
IL8RB	Cmkar2	CXCR2, CMKAR2, IL8RB, IL8R2
IL8RBP	-	IL8RP (pseudogene)
GPR9	Cmkar3	CXCR3, CMKAR3, GPR9, CKR-L2, IP10/Mig-R, IP10-R
CXCR4	Cmkar4	CMKAR4, LCR1, NPY3R, fusin, HM89, LESTR, NPYRL, SDF-1R
BLR1	Blr1	CXCR5, BLR1, MDR15

Table from the Cytokine Family Database (http://crf.medic.kumamoto-u.ac.jp/CRF/CXCR/CXCR.html)

Table 5: C-X-C Chemokines

Gene Symbol		Gene Name
(Human)	(Mouse)	
SCYB1	Gro1	CXCL1, GRO1, GROa, MGSA-a
SCYB2	Scyb2	CXCL2, GRO2, GROb, MIP-2a, MGSA-b
SCYB3	-	CXCL3, GRO3, GROg, MIP-2b
SCYB4, SCYB4V1	-	CXCL4, CXCL4V1, PF4, PF4var1, PF4alt
(PF4-like)	-	PF4-like
SCYB5	Scyb5	CXCL5, ENA-78, LIX, AMCF-II
SCYB6	-	CXCL6, GCP-2, CKA-3

Gene Symbol		Gene Name
(GCP-2 like)	-	GCP-2 like
SCYB7	-	CXCL7, PPBP, PPBPL1, PBP, b-TG1, b-TG2, TGB1, TGB2, CTAPIII, CTAP3, NAP-2, NAP-2-L1, LA-PF4, MDGF, LDGF
(PBP-like)	-	DNA binding protein, SPBPBP
SCYB8	-	CXCL8, IL-8, MDNCF, NAP-1, 3-10C, MONAP, LUCT, AMCF-I, LYNAP, NAF, b-ENAP
SCYB9	Mig	CXCL9, mig, Humig
SCYB10	Ifi10	CXCL10, IP-10, crg-2, mob-1, C7, gIP-10
SCYB11 (SCYB9B)	-	CXCL11, H174, b-R1, I-TAC, IP-9
SCYB12	Sdf1	CXCL12, SDF-1a, SDF-1b, PBSF, TLSF-a, TLSF-b, TPAR1
SCYB13	-	CXCL13, BLC, BCA-1, BLR1L, Angie
SCYB14	Scyb14	CXCL14, BRAK, NJAC
SCYB15	Scyb15	CXCL15, lungkine, CINC-2b-like, weche
(MGSA-pseudo)	-	MGSA pseudogene
(NAP-4)	-	NAP-4

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/CFC/CK/CXCG/CXCG.html)

Surrogate ligands for orphan receptors that exhibit homology to the CXXXC family of chemokine receptors (*e.g.*, CX3CR1) can be obtained by shuffling different forms of nucleic acids that encode SCYD-1 (*e.g.*, homologs of SCYD-1 from different mammalian species). Similarly, surrogate ligands for C chemokine-like receptors (*e.g.*, CCXCR1 (gene names include Ccxcr1, XCR1, GPR5, SCM1-R) can be obtained by shuffling nucleic acids that encode known C chemokines, such as those listed in Table 6.

Table 6: C Chemokines

Gene Symbol		Gene Name
(Human)	(Mouse)	
SCYC1	Lptn	CL1, Lymphotactin, SCM-1a, ATAC
SCYC2	-	CL2, SCM-1b

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/CFC/CK/CG/CG.html)

Chemokines that are encoded by viruses are also of interest for use in obtaining surrogate ligands for orphan receptors. For example, one can shuffle two or more viral chemokine-encoding nucleic acids listed in Table 7.

Table 7: Viral Chemokine cDNAs and Corresponding GenBank Accession Numbers

Marek's disease virus (Gallid herpesvirus 1)
M89471 Eco Q protein
U34965 Eco Q protein
U34966 Eco Q protein
U55025 MKT-1 unidentified

Stealth virus (unclassified)
AF145588 clone 3B516
U27769 clone 3B654 M13RP
U27885 clone 3B33 T7
U27908 clone 3B624 T7
U27928 clone 3B657 T7

Kaposi's sarcoma-associated herpes virus-HHV8
U50138 vMIP-1a
U71366 similar to MIP-1a
U74585 vMIP-IA
U75698 vMIP-I
U93872 K6

Kaposi's sarcoma-associated herpes virus-HHV8
AF091347 1609-1325
U67775 vMIP-1B
U71365 similar to MIP-1a
U75698 vMIP-II
U93872 K4

Kaposi's sarcoma-associated herpes virus-HHV8
AF091347 972-628
U75698 22185-22529
U83351 BCK
U93872 K4.1

Molluscum contagiosum virus subtype 1
U60315 MC148R
U86945 H-M-N-3

Molluscum contagiosum virus subtype 2
U96749 MC148R2

Murine cytomegalovirus 1
AF124602 CC chemokine homolog
L32187 2747-2942
U10326 MCK-1 (ORF HJ1)
U68299 188029-188376

Human herpesvirus-6 variant A strain U1102
U13194 EDRF3
X83413 U83 (EDRF3)

Human herpesvirus-6 variant B strains CB11R Z29 and HST

AB021506 U83
AF157706 herpesvirus 6B
U92288 H83
Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/CFC/CK/VIRUS/VIRUS.html)

2. FGF Family

To obtain surrogate ligands for orphan receptors that exhibit homology to the fibroblast growth factor (FGF) receptor family, the invention involves shuffling two or more forms of an FGF-encoding nucleic acid. Again, one can use homologs of a single FGF species that are obtained from different mammals, or two or more types of FGF species from a single mammalian species, or a combination thereof. Genes that encode members of the FGF/HBGF family are listed in Table 8.

Table 8: FGF/HBGF Family

Gene Symbol (Human)	(Mouse)	Gene Name
FGF1	Fgf1	fibroblast growth factor 1 (acidic), acidic FGF, heparin-binding growth factor-1 (HBGF-1), FGFA, beta-endothelial cell growth factor (ECGF-beta)
FGF2	Fgf2	fibroblast growth factor 2 (basic), basic FGF, heparin binding growth factor-2 (HBGF-2), bFGF
FGF3	Fgf3	fibroblast growth factor 3, int-2, (murine mammary tumor virus integration site (v-int-2) oncogene homolog)
FGF4</A< TD>	Fgf4	fibroblast growth factor 4, transforming gene from human stomach-1, hst, hst-1, heparin-binding secretary transforming factor-1 (HSTF1), Kaposi's sarcoma FGF (ksFGF), K-FGF, KS3
FGF5	Fgf5	fibroblast growth factor 5, oncogene encoding fibroblast growth factor-related protein
FGF6	Fgf6	fibroblast growth factor 6, fibroblast growth factor-related gene, hst-2
FGF7	Fgf7	fibroblast growth factor 7, keratinocyte growth factor (KGF)
FGF8	Fgf8	fibroblast growth factor 8, androgen-induced growth factor (AIGF)
FGF9	Fgf9	fibroblast growth factor 9, glia-activating factor (GAF), FGF-9
FGF10	Fgf10	fibroblast growth factor 10, keratinocyte growth factor 2, KGF-2
FGF11	Fgf11	fibroblast growth factor 11, fibroblast growth factor homologous factor 3 (FHF-3)

Gene Symbol (Human)	(Mouse)	Gene Name
FGF12	Fgf12	fibroblast growth factor 12, fibroblast growth factor homologous factor 1 (FHF-1)
FGF13	Fgf13	fibroblast growth factor 13, fibroblast growth factor homologous factor 2 (FHF-2)
FGF14	Fgf14	fibroblast growth factor 14, fibroblast growth factor homologous factor 4 (FHF-4)
(FGF15)	Fgf15	fibroblast growth factor 15
FGF16	-	fibroblast growth factor 16
FGF17	Fgf17	fibroblast growth factor 17
FGF18	Fgf18	fibroblast growth factor 18
FGF19	Fgf18	fibroblast growth factor 19
(FGF20)	-	XFGF-20
(FGF21)	-	fibroblast growth factor 21
(FGFH)	-	fibroblast growth factor homologous
(C05D11.4.)	-	hypothetical 48.1 KD protein COD11.4

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/)

3. IL-6 Family

Nucleic acids that encode members of the IL-6 family can be shuffled to obtain surrogate ligands for orphan receptors that exhibit homology to the IL-6 receptor family. Suitable nucleic acids that encode members of the IL-6 family include those listed in Table 9.

Table 9: IL-6 Family

Gene Symbol (Human)	(Mouse)	Gene Name
IL6	Il6	interleukin 6, B-cell stimulatory factor-2 (BSF-2), interferon-beta 2
CSF3	Csfg	colony stimulating factor 3, granulocyte colony stimulating factor
(MGF)	-	myelomonocytic growth factor

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/)

4. LIF/OSM Family

Similarly, nucleic acids that encode members of the leukemia inhibitory factor/oncostatin M family of ligands can be shuffled to obtain surrogate ligands for orphan receptors that exhibit homology to a known member of the LIF/OSM receptor family. Nucleic acids that encode LIF/OSM ligands include those listed in Table 10.

Table 10: LIF/OSM Family

Gene Symbol (Human)	Gene Symbol (Mouse)	Gene Name
LIF	Lif	leukemia inhibitory factor, cholinergic differentiation factor
OSM	Osm	oncostatin M

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/)

5. MDK/PTN Family

To obtain surrogate ligands for orphan receptors that exhibit homology to receptors for the MDK/PTN family of cytokines, one can shuffle nucleic acids that encode one or more of these cytokines. Representative examples are shown in Table 11.

Table 11: MDK/PTN Family

Gene Symbol (Human)	Gene Symbol (Mouse)	Gene Name
MDK	mdk	midkine, retinoic acid-induced heparin-binding protein (RI-HB), neurite growth-promoting factor-2 (NEGF2), retinoic acid-responsive protein
-	Mdk-ps1	midkine pseudogene 1
PTN	ptn	pleiotrophin (PTN), heparin-binding neutrophic factor (HBNF-1), osteoblast specific protein (OSF-1), heparin-binding growth factor 8 (HBGF-8), heparin-binding growth-associated molecule (HB-GAM), neurite growth-promoting factor-1 (NEGF1, osteoblast stimulating factor-1)

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/)

6. NGF Family

Nucleic acids that encode members of the nerve growth factor (NGF) family can be shuffled to obtain surrogate ligands for orphan receptors that exhibit homology to the

NGF receptor family. Suitable nucleic acids that encode members of the NGF family include those listed in Table 12.

Table 12: NGF Family

Gene Symbol (Human)	Gene Symbol (Mouse)	Gene Name
BDNF	Bdnf	brain-derived neurotrophic factor
NGFB	Ngfb	Nerve growth factor, beta NGF
NTF3	Ntf3	neurotrophin-3, NT-3, NGF-2
NTF5	Ntf5	neurotrophin-4, neurotrophin-5, NT-4, NT-5
NTF6A	-	neurotrophin-6 alpha, NT-6 alpha
NTF6B	-	neurotrophin-6 beta, NT-6 beta
NTF6G	-	neurotrophin-6 gamma, NT-6 gamma
(NTF7)	-	neurotrophin-7
Unclassified	-	-

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/)

7. TNF Family

Nucleic acids that encode members of the tumor necrosis factor (TNF) family can be shuffled to obtain surrogate ligands for orphan receptors that exhibit homology to the TNF receptor family. Suitable nucleic acids that encode members of the TNF family include those listed in Table 13.

Table 13: TNF Family

Gene Symbol (Human)	Gene Symbol (Mouse)	Gene Name
TNF	Tnf	tumor necrosis factor, TNFα (Tumor Necrosis Factor α), TNF superfamily member 2 (TNFSF2)
LTA	Lta	Lymphotoxin, Lymphotoxin α, TNF superfamily member 1 (TNFSF1), TNFβ
LTB	Ltb	Lymphotoxin β, TNF superfamily member 3 (TNFSF3), TNFC
TNFSF3L	Tnfsf3l	TNF superfamily member 3 (LTB)-like peptidoglycan recognition protein, peptidoglycan recognition protein precursor (PGRP)
TNFSF4	Txgp1l	tumor necrosis factor ligand superfamily member 4, TXGP1, OX-40 ligand, tax-transcriptionally activated glycoprotein 1 ligand
TNFSF5	Tnfsf5	tumor necrosis factor ligand superfamily member 5, CD40 antigen

Gene Symbol (Human)	(Mouse)	Gene Name
		ligand, CD40LG, CD40L, TNF-related activation protein (TRAP), hyper-IgM syndrome, gp39
TNFSF6	Fasl	tumor necrosis factor ligand superfamily member 6, apoptosis (APO-1) antigen ligand 1, APT1LG1, Fas ligand (FASL)
TNFSF7	Tnfsf7	tumor necrosis factor ligand superfamily member 7, CD70 antigen, CD70, CD27 ligand, CD27LG, CD27L
TNFSF8	Tnfsf8	tumor necrosis factor ligand superfamily member 8, CD30 antigen ligand, CD30LG, CD30L
TNFSF9	Tnfsf9	tumor necrosis factor ligand superfamily member 9, 4-1BB ligand, 4-1BBLG, CD antigen 137 ligand
TNFSF10	Trail	tumor necrosis factor ligand superfamily member 10, Apoptosis ligand TRAIL, Apo-2 ligand, TNF-RELATED APOPTOSIS INDUCING LIGAND (TRAIL), TL2
TNFSF11	Tnfsf11	tumor necrosis factor ligand superfamily member 11, TNF-related activation-induced cytokine receptor activator of nuclear factor kappa B ligand (RANKL), osteoprotegerin ligand, TNF-related ligand (TRANCE), ODF
TNFSF12	Tnfsf12	tumor necrosis factor ligand superfamily member 12, TNF-related weak inducer of apoptosis, TWEAK
TNFSF13	-	tumor necrosis factor ligand superfamily member 13
TNFSF14	-	tumor necrosis factor ligand superfamily member 14, LIGHT, lymphotoxin-beta receptor (LTbR), ligand for herpesvirus entry mediator (HVEML)
TNFSF15	-	tumor necrosis factor ligand superfamily member 15, TL1
TNFSF18	-	tumor necrosis factor ligand superfamily member 18 (TNFSF18), AIRTL, GITRL, glucocorticoid-induced TNFR-related protein ligand (TNFSF18), AITR ligand (TL6)
TNFSF19	Tnfsf19-pending	tumor necrosis factor ligand superfamily member 19, KE05 protein, FLDED-1, death effector domain-containing protein (DEDD)

Table from the Cytokine Family Database (http://cytokine.medic.kumamoto-u.ac.jp/)

8. TGF-β Family

Nucleic acids that encode members of the transforming factor-β (TGF-β) family can be shuffled to obtain surrogate ligands for orphan receptors that exhibit homology to the TGFβ receptor family. Suitable nucleic acids that encode members of the TGFβ family include those listed in Table 14.

Table 14: TGFβ Family

Mullerian inhibitory substance (MIS)
Inhibins
Bone morphogenetic proteins [4] BMP-2, BMP-3 (osteogenin), BMP-3B (GDF-10),
BMP-4 (BMP-2B), BMP-5, BMP-6 (VGR-1), BMP-7 (OP-1) and BMP-8 (OP-2)
Embryonic growth factor GDF-1
Growth/development factor GDF-5
Growth/development factor GDF-3, GDF-6, GDF-7, GDF-8 (myostatin) and GDF-9
Mouse protein nodal
Chicken dorsalin-1 (dsl-1)
Xenopus vegetal hemisphere protein Vg1
Drosophila decapentaplegic protein (DPP-C)
Drosophila protein screw (scw)
Drosophila protein 60A
Caenorhabditis elegans larval development regulatory growth factor daf-7
Mammalian endometrial bleeding-associated factor (EBAF)
Mammalian glial cell line-derived neurotrophic factor (GDNF)

Once the nucleic acids are shuffled, the gene products of the shuffled nucleic acids are screened to identify those that exhibit the desired activity on the orphan receptor.

B. Screening Methods

A recombination cycle is usually followed by at least one cycle of screening or selection for molecules having a desired property or characteristic. For example, a library of recombinant polynucleotides can be screened to identify those that encode a polypeptide that can act as a surrogate ligand for an orphan receptor.

1. General considerations.

If a recombination cycle is performed *in vitro*, the products of recombination, *i.e.*, recombinant segments, are sometimes introduced into cells before the screening step. Recombinant segments can also be linked to an appropriate vector or other regulatory sequences before screening. Alternatively, products of recombination generated *in vitro* are sometimes packaged as viruses before screening. If recombination is performed *in vivo*, recombination products can sometimes be screened in the cells in which recombination occurred. In other applications, recombinant segments are extracted from the cells, and optionally packaged as viruses, before screening.

The nature of screening or selection depends on what property or characteristic is to be acquired or the property or characteristic for which improvement is sought, and several examples are discussed below. It is not usually necessary to understand the molecular basis by which particular products of recombination (recombinant segments) have acquired new or improved properties or characteristics relative to the starting substrates. Screening/selection can then be performed, for example, for recombinant surrogate ligands that have increased agonist activity on a target cell that displays the receptor of interest without the need to attribute such improvement to any of the individual component sequences of the surrogate ligand.

Depending on the particular screening protocol used for a desired property, initial round(s) of screening can sometimes be performed in bacterial cells due to high transfection efficiencies and ease of culture. Later rounds, and other types of screening which are not amenable to screening in bacterial cells, are performed in mammalian cells to optimize recombinant segments for use in an environment close to that of their intended use. Final rounds of screening can be performed in the precise cell type of intended use (*e.g.*, a human cell).

The screening or selection step identifies a subpopulation of recombinant polynucleotides that encode polypeptides that have evolved toward acquisition of a new or improved desired receptor binding and/or modulatory activity. Depending on the screen, the recombinant polynucleotides can be identified as components of cells, components of viruses or in free form. More than one round of screening or selection can be performed after each round of recombination.

If further improvement in a property is desired, at least one and usually a collection of recombinant polynucleotides surviving a first round of screening/selection are subject to a further round of recombination. These recombinant polynucleotides can be recombined with each other or with exogenous segments representing the original substrates or further variants thereof. Again, recombination can proceed *in vitro* or *in vivo*. If the previous screening step identifies desired recombinant polynucleotides as components of cells, the components can be subjected to further recombination *in vivo*, or can be subjected to further recombination *in vitro*, or can be isolated before performing a round of *in vitro* recombination. Conversely, if the previous screening step identifies desired recombinant polynucleotides in naked form or as components of viruses, these polynucleotides can be introduced into cells to perform a round of *in vivo* recombination. The second round of recombination, irrespective how performed, generates further recombinant polynucleotides which encompass additional diversity than is present in recombinant segments resulting from previous rounds.

The second round of recombination can be followed by a further round of screening/selection according to the principles discussed above for the first round. The stringency of screening/selection can be increased between rounds. Also, the nature of the screen and the property being screened for can vary between rounds if improvement in more than one property is desired or if acquiring more than one new property is desired. Additional rounds of recombination and screening can then be performed until the recombinant segments have sufficiently evolved to acquire the desired new or improved property or function.

Various screening methods for particular applications are described herein. In some instances, screening involves expressing the recombinant peptides or polypeptides encoded by the recombinant polynucleotides of the library as fusions with a protein that is displayed on the surface of a replicable genetic package. For example, phage display can be used. *See, e.g*, Cwirla *et al*., *Proc. Natl. Acad. Sci.* USA 87: 6378-6382 (1990); Devlin *et al*., *Science* 249: 404-406 (1990), Scott & Smith, *Science* 249: 386-388 (1990); Ladner *et al*., US 5,571,698. Other replicable genetic packages include, for example, bacteria, eukaryotic viruses, yeast, and spores.

The genetic packages most frequently used for display libraries are bacteriophage, particularly filamentous phage, and especially phage M13, Fd and F1. Most work has involved inserting libraries encoding polypeptides to be displayed into either gIII or gVIII of these phage forming a fusion protein. *See, e.g.*, Dower, WO 91/19818; Devlin, WO 91/18989; MacCafferty, WO 92/01047 (gene III); Huse, WO 92/06204; Kang, WO 92/18619 (gene VIII). Such a fusion protein comprises a signal sequence, usually but not necessarily, from the phage coat protein, a polypeptide to be displayed and either the gene III or gene VIII protein or a fragment thereof. Exogenous coding sequences are often inserted at or near the N-terminus of gene III or gene VIII although other insertion sites are possible.

Eukaryotic viruses can be used to display polypeptides in an analogous manner. For example, display of human heregulin fused to gp70 of Moloney murine leukemia virus has been reported by Han *et al.*, *Proc. Natl. Acad. Sci. USA* 92: 9747-9751 (1995). Spores can also be used as replicable genetic packages. In this case, polypeptides are displayed from the outer surface of the spore. For example, spores from *B. subtilis* have been reported to be suitable. Sequences of coat proteins of these spores are provided by Donovan *et al.*, *J. Mol. Biol.* 196, 1-10 (1987). Cells can also be used as replicable genetic packages. Polypeptides to be displayed are inserted into a gene encoding a cell protein that is expressed on the cells surface. Bacterial cells including *Salmonella typhimurium*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Klebsiella pneumonia*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Bacteroides nodosus*, *Moraxella bovis*, and especially *Escherichia coli* are preferred. Details of outer surface proteins are discussed by Ladner et al., US 5,571,698 and references cited therein. For example, the *lam*B protein of *E. coli* is suitable.

A basic concept of display methods that use phage or other replicable genetic package is the establishment of a physical association between DNA encoding a polypeptide to be screened and the polypeptide. This physical association is provided by the replicable genetic package, which displays a polypeptide as part of a capsid enclosing the genome of the phage or other package, wherein the polypeptide is encoded by the genome. The establishment of a physical association between polypeptides and their genetic material allows simultaneous mass screening of very large numbers of phage bearing different

polypeptides. Phage displaying a polypeptide with affinity to a target, *e.g.*, a receptor, bind to the target and these phage are enriched by affinity screening to the target. The identity of polypeptides displayed from these phage can be determined from their respective genomes. Using these methods a polypeptide identified as having a binding affinity for a desired target can then be synthesized in bulk by conventional means.

2. Screening assays for surrogate ligand or surrogate receptor activity

Screening of the recombinant libraries can involve identifying those members that encode a polypeptide that specifically binds to the receptor of interest. The libraries of recombinant polynucleotides are expressed and those that can bind to the receptor with a desired specificity and avidity are chosen for use, or for further improvement. In presently preferred embodiments, the library of recombinant polypeptides are displayed on the surface of a replicable genetic package.

For some applications, a binding assay is sufficient to identify a surrogate ligand or surrogate receptor. However, in other applications, it is desirable to obtain a surrogate that exerts a biological activity upon binding to its orphan counterpart. The biological activity assay can be conducted after pre-screening using a binding assay, or can be used on its own without a prescreen.

In some embodiments, the libraries of recombinant polynucleotides are screened by expressing the library and contacting the resulting library of candidate surrogate ligands with a test cell that contains the receptor of interest, or at least a sufficient portion for biological activity. Suitable test cells are those that are known to allow biological activity for previously known members of the ligand family to which the surrogate ligand presumably belongs.

For receptors such as cytokine receptors, the extracellular domain of the receptor of interest is expressed as a fusion with the cytoplasmic domain of a known receptor. The transmembrane domain of the known receptor or of the receptor of interest can also be included in the fusion protein. The fusion protein is displayed on a cell that is permissive for the biological activity of known ligands for the receptor family to which the receptor of interest is presumed to belong. Upon binding of a surrogate ligand to the extracellular domain, the biological activity is observed.

In some embodiments, the screening methods of the invention use a cell that contains a polypeptide that has a ligand binding domain of the receptor of interest (*e.g.*, an orphan receptor). The polypeptide will also include a DNA binding domain, which can be that of the orphan receptor, or more preferably is obtained from a known receptor or is a DNA binding domain for which the response element is known (*e.g.*, Gal4, nuclear hormone receptors, and the like). Examples of suitable chimeric polypeptides are described in more detail above. Conveniently, the chimeric receptor polypeptide is introduced into the cell by expression of a polynucleotide that encodes the receptor polypeptide. For example, an expression vector that encodes the chimeric receptor can be introduced into the cell that is to be used in the assay.

For a nuclear receptor, the cells preferably also contain a response element that can be bound by the DNA binding domain. The response element is operably linked to a promoter that is active in the cell. In presently preferred embodiments, the promoter is operably linked to a reporter gene that, when expressed, produces a readily detectable product. The response element/reporter gene construct is conveniently introduced into cells as part of a "reporter plasmid."

For some screening assays, it is desirable to present to the assay a standard amount of the ligand being tested. In such instances, one can "tail" the ligands with a suitable affinity tag and express the ligands in an expression system known to allow biological activity for the previously known members of the family to which the ligand presumably belongs. Cell extracts and/or supernatants that contain the expressed ligands can be simultaneously affinity purified in a batchwise fashion, for example, in pools, and eluted. The system can be calibrated such that differences in expression level of the different ligands (which differences are likely to occur) would not result in differences in the total amount of ligand presented in an assay. For example, one can use 10-50-fold excess ligand over the capacity of the affinity purification support.

In assays in which pools are processed, the levels of individual members within each pool will not be identical. In such situations, positive pools are identified without concern for false negatives due to poor expression of any particular ligand surrogate.

3. Screening assays to identify compounds that modulate activity of a surrogate ligand or surrogate receptor.

The invention also provides screening assays for identifying compounds that can modulate the biological activity of a surrogate ligand or a surrogate receptor obtained using the methods of the invention. These compounds can function by, for example, altering the interaction between the receptors and their ligands, or between the receptors and the remainder of the signal transduction pathway. Compounds that are identified using the screening methods of the invention find use in studies of interactions between the ligand and receptor and in studies of signal transduction. The compounds also find therapeutic use in situations in which it is desirable to increase or decrease expression of genes that are under the control of a particular receptor. Other uses will also be apparent those of ordinary skill in the art.

In the screening methods for obtaining modulators, a test system such as those described above can be used. For example, host cells that contain a reporter plasmid, a chimeric receptor polypeptide, and the surrogate ligand are incubated in the presence of a test compound. Essentially any chemical compound can be used as a potential modulator in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (*e.g.*, in microtiter formats on microtiter plates in robotic assays). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs Switzerland) and the like.

In one preferred embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential therapeutic compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*, the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.*, U.S. Patent 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (PCT Publication No. WO 91/19735), encoded peptides (PCT Publication WO 93/20242), random bio-oligomers (PCT Publication No. WO 92/00091), benzodiazepines (U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with β-D-glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see*, Ausubel, Berger and Sambrook, *all supra*), peptide nucleic acid libraries (*see, e.g.*, U.S. Patent 5,539,083), antibody libraries (*see, e.g.*, Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.*, Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent 5,593,853), small organic molecule libraries (*see, e.g.*, benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337; benzodiazepines, 5,288,514, and the like).

Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.*, 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.*, ComGenex, Princeton, N.J., Asinex, Moscow, Ru, Tripos, Inc., St. Louis, MO, ChemStar, Ltd, Moscow, RU, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, *etc.*).

EXAMPLES

The following examples are offered to illustrate, but not to limit the present invention.

The following abbreviations are used herein: IFN-α, alpha interferon; Hu-IFN-α, human IFN-α; Mu-IFN-α, murine IFN-α; HTP, high throughput; CHO, Chinese hamster ovary; EPO, erythropoietin; GM-CSF, granulocyte macrophage colony stimulating factor; G-CSF, granulocyte colony stimulating factor; IL, interleukin; PBS, phosphate buffered saline; CPE, cytopathic effect.

Example 1

RAPID EVOLUTION OF A CYTOKINE USING MOLECULAR BREEDING

Molecular breeding is the application of classical breeding to sub-genomic sequences. This approach to sequence evolution generalizes concepts from classical genetics, allowing one to selectively breed DNA sequences in the test tube. In this study, *in vitro* DNA shuffling was used to breed a family of over 20 human interferon alpha (Hu-IFN-α) genes for increased anti-viral and anti-proliferation activities in murine cells. Only 68 assays of pools of interferons were used to obtain a clone with 135,000-fold improved specific activity over Hu-IFN-α2a in the first cycle of shuffling. After a second cycle of selective breeding, the most active clone was improved 285,000 relative to Hu-IFN–α2a. Remarkably, the three most active clones are more active than the native murine IFN-αs. These chimeras are derived from up to five parental genes, but contain no random point mutations. These results demonstrate that diverse cytokine gene families can be used as breeding stock from which to rapidly evolve cytokines that are more active or have superior

selectivity profiles than native cytokine genes. Molecular breeding provides an economical alternative to genomics-based approaches to searching for potent activities of interest in existing genomes.

Introduction

Alpha interferons are members of the diverse helical-bundle super-family of cytokine genes that contains many clinically important pharmaceutical proteins such as EPO, GM-CSF, G-CSF, IFN-α, IFN-β, IL-2, IL-3, IL-4 and several other interleukins (Sprang and Bazan (1993) *Current Opinion in Structural Biology* 3:815-827). While these proteins have important therapeutic value in the treatment of a number of diseases, they have not been optimized by natural selection as pharmaceuticals. For example, dose-limiting toxicity, receptor cross-reactivity, and short serum half-lives significantly reduce the clinical utility of many of these cytokines (Dusheiko, G. (1997) *Hepatology* **26(3 Suppl 1):**112S-121S; Vial and Descotes (1994) *Drug Experience* **10 (2):** 115-150; Funke *et al.* (1994) *Ann. Hematol.* **68(1):**49-52; Schomberg *et al.* (1993) *J. Cancer Res. Clin. Oncol.* **119(12):**745-55). Molecular breeding provides a general method for improving these properties.

The cytokine super-family has evolved by a series of gene duplications and recombination events. For example, the α, β and ω interferons are derived by ancient duplication of a common ancestor with subsequent recombination within the IFN-α gene family (Hughes, A. L. (1995) *J. Mol. Evol.* **41(5):** 539-48). Similarly, the genes encoding IL-4 and IL-13 are in proximity in human and murine genomes and they share several, but not all, of their biological functions (Punnonen *et al.* (1993) *Proc. Nat'l. Acad. Sci. USA* **90(8):**3730-4), suggesting that they have arisen by gene duplication. The receptors for the cytokine supergene family have also been generated by duplication, mutation, and recombination of a few modular receptor domains (Uze *et al.* (1995) *J. Interferon Cytokine Res.* **15(1):**3-26; Bazan *et al.* (1990) *Proc. Nat'l. Acad. Sci. USA* **87(18):**6934-8).

The human IFN-αs are encoded by a family of over twenty tandemly duplicated non-allelic genes that share 85-98% sequence identity at the amino acid level (Henco *et al.* (1995) *J. Mol. Biol.* **185(2):**227-60). These proteins have potent antiviral and anti-proliferative activities that have great clinical utility as anticancer and antiviral therapeutics. While the utility of chimeric IFNs derived from this gene family has been

recognized (Horisberger and Di Marco (1995) *Pharmacol. Ther.* **66(3):**507-34), only a small fraction of the 10^{26} possible chimeras have been explored either in natural human evolution or by the methods of modern molecular biology; and only one natural IFN-α subtype, Hu-IFN-α2, has been used in extensive clinical studies (*Id.*). The most active engineered IFN-α, IFN-Con_1, is a consensus of thirteen wild type Hu-IFN-α genes that is currently being used in hepatitis C therapy (Blatt *et al.* (1996) *J. Interferon Cytokine Res.* **16(7):**489-99).

DNA shuffling, or molecular breeding, is a method for permutation of natural genetic diversity. This technology provides a powerful tool for rapidly evolving single genes, operons and whole viruses for desired properties (Stemmer, W. P. C. (1995) *Biotechnology* **13**: 549-555; Patten *et al.* (1996) *Current Opinion in Biotechnology* **8**:724-733; Crameri *et al.* (1998) *Nature* **15**:288-91), and has many advantages relative to random mutation or rational sequence design. This Example describes the use of family DNA shuffling to rapidly evolve the Hu-IFN-α gene family for activity in mouse cells. The native Hu-IFN-α genes are 53-65% identical to Mu-IFN-αs and exhibit very weak activity on murine cells (Horisberger and Di Marco, *supra.*). Similarly, the extra-cellular domains of the IFN-α receptors share only 49% sequence identity (Uze *et al.*, *supra.*). Despite these sequence differences, we obtained shuffled IFN-αs that are more potent in mouse cells than the native Mu-IFN-αs.

Experimental protocols

DNA cloning, sequencing and shuffling

The Hu-IFN-α gene family was PCR amplified from human genomic DNA using twelve sets of degenerate primers. Three hundred micrograms of PCR product was fragmented with DNase I, 25 - 60 bp fragments were gel purified, and family shuffling of the fragments was performed as described (Crameri *et al.* (1998) *Nature* **15**:288-91). Two additional libraries of shuffled Hu-IFN-α genes we. made from eight cloned Hu-IFN genes (Hu-IFN-αs 1, 4, 5, 6, 14, 16, 17 and F). Fragments of 25-50 or 50-100 bp were purified, and shuffling was done as described (Crameri *et al.*, *supra.*). Assembled insert was cloned by standard methods into the phagemid display vector pDEI-932. Hu-IFN-α-Con1 was constructed from synthetic oligonucleotides. Hu-IFN-αs 1, 2a, 4, 5, 6, 14, 16, 17 and F; and

Mu-IFN-αs 1, 4 and 6 were cloned from genomic DNA and sequenced on an ABI DNA sequencer.

DNA sequence analysis

The extracellular domains of the human and mouse IFN-α receptors were aligned by the Clustal method (DNA STAR; SWISS-PROT accession numbers P33896, P17181, P48551; GENBANK accession number AF013274).

Phagemid display of IFN

For HTP primary screening of activity, shuffled Hu-IFN-α genes were expressed in a biologically active form by phage display, similarly to the expression strategy used for other four helix bundle cytokines. The phagemid display vector pDEI-932 is a standard gene III phagemid display vector wherein the STII leader is fused to the amino terminus of Hu-IFN-α and the E-tag (Pharmacia) plus a 6-His tag is fused to the carboxyl terminus. Immediately following the C-terminal tag is a suppressible amber codon, followed by M13 gene III (fused at residue 247 of gene III). The IFN-α gene III insert is under the control of the pBAD promoter, and the backbone plasmid is an Amp^R derivative of pBR322 containing an M13 origin of replication. Large scale (250 ml) phagemid preps were done by standard methods (Klaus *et al.* (1997) *J. Mol. Biol.* **274(4)**:661-75) in the presence of 0.002% arabinose to induce expression of the IFN-α gene III fusion. Phagemids were PEG precipitated, CsCl banded, and dialyzed into PBS prior to assaying.

HTP phagemid preparations

For the purposes of HTP screening, *E. coli* harboring phagemids were picked with a Q-BOT robotic colony picker (Genetix) into 96-well plates containing 100 microliters of 2XYT per well. Confluent cultures were grown overnight at 37° C. The overnight cultures were diluted 20-fold into fresh 2XYT, Amp/0.002% arabinose/10^{10} pfu/ml M13 VCS helper phage and grown for four hours with vigorous shaking. The cells were pelleted and phage supernatants were transferred to 96-well dialysis plates containing a 100 kilodalton cutoff membrane prior to assaying. Samples were dialyzed against PBS and then filter sterilized through 96-well 0.45 micron membranes. Sterile phagemid samples were used directly in cellular assays.

Antiviral assays

Antiviral activities were determined by the cytopathic effect (CPE) reduction assay on mouse L929 cells challenged with encephalomyocarditis virus (EMCV). Briefly, target cells were grown to confluence, trypsinized, and distributed into 96 well flat bottom microtitre plates (10^4 cells per well) in RPMI medium supplement with 10%FCS and Penicillin/Streptomycin antibiotics. IFN-α samples were titrated in triplicate in 5 fold dilutions. After incubation for 16 hours, the medium was removed, replaced with medium containing EMCV (100 $TCID_{50}$ per well) and the plates were incubated for 2 days until CPE occurred. Medium was removed, the cells were washed twice with PBS, and neutral red (1:100 dilution) was added and incubated for 2 hours. During the last 20 minutes, cells were fixed with 0.5% glutaraldehyde. The unstained dye solution was removed, the plates were washed twice with PBS, and the color was extracted with 50% methanol, 1% acetic acid. The extracted dye solution in the well was quantitated colorimetrically at 540 nanometers with a spectrophotometer. Results of the CPE reduction assay derived as above were plotted to produce sigmoidal dose-response curves by plotting the logarithm of the IFN-α concentration versus the cell viability. One unit/ml is defined as the interpolated IFN-α concentration giving 50% protection (on a scale of 0 to 100% determined by controls with no IFN-α and with or without virus).

Deconvolution of libraries

In cycle one, eight pools of 12 were assayed, and one had measurable antiviral activity. Sixteen pools of 96 were assayed, the most active pool of 96 was broken into eight pools of 12, and these pools were assayed separately. Three pools of 12 had measurable activity, and thirty-six individual phagemids were prepared, purified and assayed from these pools. One chimera (Hu-IFN-α-CH1.4) was obtained by randomly screening individual clones in the library for L929 antiviral activity. Three IFN-α phagemids with antiviral activity were obtained (one from each pool). The IFN-α chimeras from these phagemids were cloned into the CHO expression vector pDEI-1011, transfected, and purified as described.

Construction of round two libraries

Five cycle two libraries were constructed by shuffling equimolar qualities of plasmid DNA In the following combinations: CH1.1 x CH1.2; CH 1.1 x CH1.3; CH1.1 x CH1.4; CH 1.2 x CH 1.3; CH 1.1 x CH 1.2 x CH 1.3 x CH 1.4. Shuffled libraries were made in pDEI-932 from 25-50 bp fragment assemblies as described (Crameri *et al.*, *supra.*).

HTP proliferation assays

The L929 anti-proliferative assay was performed according to standard ^{3}H thymidine incorporation methods. Briefly, IFN-α samples were titrated in triplicate in 5 fold dilution steps down the plate. For HTP screening in the second round of shuffling, four single 10-fold dilutions were assayed in the primary screen, and subsequent rescreens were done in triplicate. L929 cells (1000/well) were incubated for 72 hours at 37° C, 5% CO2 incubator. During the last 16 hours of incubation, 1 μCi/well of ^{3}H thymidine was added. The plates were then harvested on a Harvester-96 (Tomtec) and thymidine incorporation was counted on a beta counter (Microbeta, Wallac).

CHO expression and purification of shuffled IFN-αs

IFN-α genes were cloned into a standard CHO expression vector (pDEI-1011) in which the E-tag/6-His tag (Pharmacia) is fused to the C-terminus of the IFN-αs. Expression is driven by the SR-α promoter, and stable transfectants were selected at 1 mg/ml G418. The four most active clones from the first round and the fifteen most active clones from the second round were inserted into a pDEI-1011, introduced into CHO cells by transfection (Sambrook *et al.*, *supra.*), and the proteins were affinity purified from the supernatant on anti-E tag Sepharose (Pharmacia).

Daudi proliferation assays

Eight chimeric phage-displayed IFN-αs were sequenced from randomly picked clones. Four of the eight sequences encoded in-frame IFN-α genes. These four chimeras and Hu-IFN-α2a were expressed, purified, and assayed for anti-proliferation activity on human Daudi cells. The Daudi antiproliferation assay was done as described (Scarozza *et al.* (1992) *J. Interferon Res.* **12**: 35-42). One unit/ml is defined as the concentration giving half-maximal inhibition of proliferation. Two thirds of the clones in the

cycle two library were more potent than Hu-IFN-α-CH1.4 in the HTP L929 antiproliferation assay.

Results And Discussion

Two rounds of molecular breeding and screening were performed. In the first round, family DNA shuffling by homologous *in vitro* recombination was used to make a library of chimeric Hu-IFN-αs. All of the Hu-IFN-α genes, including pseudogenes, were shuffled in order to capture the diversity of the entire family. Chimeric IFN-αs were expressed, purified, and screened for L929 antiviral activity as pools of 12 or 96. The active pools were deconvoluted into sequentially smaller pools until single active clones were identified. Because a pooling strategy was used, a total of only 68 murine antiviral assays was used to screen this library of 1672 clones. The most active chimeric IFN-α from round one (IFN-α-CH1.1) is derived from six parental Hu-IFN-α gene segments (Figure 1A), and is 87-fold more active than Hu-IFN-α1, the wild type Hu-IFN-α that is most active in murine cells (Table 15). The large improvement in activity that was obtained in the first round of screening of this shuffled library using only 68 assays has important implications for the range of applications of molecular breeding, as discussed below.

Table 15

Activities of Parental and Evolved IFN-αs in Murine Cells

DNA Shuffling Cycle	IFN-α Gene	Genealogy	L129 antiviral activity Units/mg x 10^{-6}	Fold Improvement In Activity vs Hu-IFN-α2a
0	IFN-α-Con$_1$	Synthetic	1.0	NA
0	Hu-IFN-α2a	Native	0.00194	NA
0	Hu-IFN-α1	Native	3.0	NA
0	Mu-IFN -α1	Native	140	NA
0	Mu-IFN- α 6	Native	116	NA
0	Mu-IFN- α 4	Native	160	NA
1	IFN-αCH1.4	IFN 1,H, F, 17, J	0.15	77
1	IFN-αCH1.3	IFN 1,5	16	8,247
1	IFN-αCH1.2	IFN 1,F	42	21,649
1	IFN-αCH1.1	IFN 1, 5, 8, 14, F	262	135,000
2	IFN-αCH2.3	CH1.1 x CH1.3	268	138,000
2	IFN-αCH2.2	CH1.1 x CH1.3	400	206,000
2	IFN-αCH2.1	CH1.1 x CH1.3	554	285,000

DNA shuffling allows one to use analogs of classical breeding methods and to extend breeding in non-classical ways such as by breeding of more than two parental genes in a single molecular breeding reaction or breeding of genes from different species (Stemmer, W. P. C. (1995) *Biotechnology* **13**: 549-555; Patten *et al.* (1996) *Current Opinion in Biotechnology* **8**:724-733; Crameri *et al.*, *supra*.). As with classical breeding, the sampling of shuffled libraries is generally non-exhaustive. Indeed, the power of breeding is that large improvements in phenotype can be achieved by recursively screening only a small subset of all theoretically possible progeny (Burbank, L., "Short-cuts into the centuries to come: better plants secured by hurrying evolution," In *Luther Burbank His Methods and Discoveries: Their Practical Application*, Vol. 1, pp. 176-210 (Whitson and Williams, eds; New York: Luther Burbank Press, 1914; Haldane, J. B. S. (1924) *Cambridge Phil. Soc. Trans.* **23**: 19-41). It is therefore important to determine the most economical selective molecular breeding strategies.

The data from this experiment begin to address this important issue. In cycle two, we compared breeding strategies by doing pooled and pair-wise matings of the four IFN-α genes from round one to make five new libraries of chimeras. Four libraries were made by pair-wise matings of the genes and one library by pooled mating of all four genes. A HTP assay was used to screen 1056 individual clones from this panel of five libraries, and the top sixty candidates were rescreened quantitatively for antiviral activity in L929 cells. The genes from the eleven most active shuffled IFN-αs were expressed in CHO cells and purified IFN-α protein was assayed. The most active IFN-α from cycle two is improved 185-fold relative to Hu-IFN-α1 and 285,000-fold relative to Hu-IFN-α2a (Figures 2, 3). Remarkably, the activities of the three most active IFN-αs exceed the activity of the most active native mouse IFN-α, Mu-IFN-α4 (Table 15, Figure 3). The most active clones from round two came from the pair-wise matings of highly active clones (Hu-IFN-α-CH1.1 x Hu-IFN-α-CH1.3), with none of the most active clones in round two coming from the pooled mating (Table 15). The superior performance of pair-wise matings relative to pooled matings may reflect sparse sampling of a population with a significantly lower average level of biological activity in clones derived from the pooled mating, due to breaking up favorable amino acid combinations such as K121 and R125, as discussed below.

Libraries of family shuffled IFN-αs have few inactive or weakly active clones. In contrast, random mutagenesis typically leads to a high frequency of gene inactivation (Muller, H.J. (1964) *Mutat. Res.* 1,2-9; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347). For example, 75% of random point mutants of residues 120–136 of Hu-IFN-α4 are inactive (Tymms *et al.* (1990) *Genet. Anal. Tech. Appl.* 7(3):53-63). To assess the knockout rate in our primary libraries, we assayed four randomly chosen intact IFN-α chimeric genes from our libraries, in a human cell proliferation assay (Daudi). All four shuffled IFN-αs are as active in human Daudi cells as is Hu-IFN-α2a, despite having 10 to 21 amino acid changes relative to the closest native Hu-IFN-α (10; Figure 1; Experimental protocols). The second round of shuffling in this study gives an additional indication of the high quality of shuffled libraries, as two thirds of the clones from the second round of shuffling are more active in mouse cells than Hu-IFN-α , the most active native Hu-IFN-α. The diversity in the libraries in this study was overwhelmingly generated by recombination of pre-existing natural sequence diversity in the gene family, with random point mutation accounting for only two sequence changes in the four round one chimeras (Figure 1A). These random mutations were removed in the second round of breeding by recombination with native gene segments, and thus there were no random point mutations in the three most active round two chimeras (Figure 1B).

The dramatic difference between family shuffled libraries and libraries made by random point mutagenesis can be understood by considering that family shuffling permutes blocks of sequence containing conservative amino acid substitutions that have been selected for function during millions of years of purifying natural selection (Stemmer, *supra.*, Patten *et al.*, *supra.*, Crameri *et al.*, *supra.*, Muller *et al.*, *supra.*). Consequently, the sequence space defined by recombination of natural diversity is highly pre-selected for function and represents an infinitesimal fraction of the sequence space accessible by random mutation. For example, the Hu-IFN-α genes differ from each other by an average of 17 residues (Henco *et al.* (1985) *J. Mol. Biol.* 185(2):227-60). There are 10^{45} 17-step random mutants of Hu-IFN-α (the number of possible recombinants of the natural Hu-IFN-α sequence diversity is as follows: the Hu-IFN-α gene family is variable at 76 sites (*Id.*). However there is a very limited range of amino acid changes at these sites. There are two,

three or four amino acid alternatives at 57, 15 and 4 sites, respectively (*Id.*), so the number of possible recombinants is $2^{57} \times 3^{15} \times 4^4 = 5 \times 10^{26}$.), whereas there are 10^{26} permutations of the natural Hu-IFN-α sequence diversity. Thus, shuffled IFN–αs sample only 10^{-19} of the random point mutant spectrum. In contrast to family shuffled libraries, an infinitesimal fraction of 17-step random point mutants of Hu-IFN-αs are expected to be active (Muller, *supra.*; Moore *et al.*, *supra.*; Tymms *et al.*, *supra.*), and these libraries of shuffled chimeras are therefore highly enriched for functional clones relative to libraries made by random point mutagenesis. This result illustrates the striking ability of family shuffling to generate progeny that differ from the parent molecules at many residues, while still retaining potent biological activity.

As a consequence of the high average activity of members of the family shuffled libraries, direct screening for biological activity is possible. The ability to directly screen for the desired biological function rather than using a surrogate screen or selection is a significant advantage over other strategies such as phage panning because one can use a small number of complex biological assays to directly obtain clones with the desired biological activity. The high quality of family shuffled libraries profoundly affects the approaches that can be taken to improving complex genetic traits. Evolution of commercially important genes and proteins may be practical even when very complex, time consuming, or expensive assays are required.

Immunogenicity is clinically significant for many recombinant pharmaceutical proteins (van der Meide and Schellekens (1997) *Biotherapy* **10(1):**39-48; Konrad, M. (1989) *Tibtech* **7**:175-179; Allegreta *et al.* (1986) *J. Clin. Immunol.* **6**:481-490). The ability to evolve proteins with immunologically conservative changes while reducing properties that impact immunogenicity such as propensity to unfold, aggregate, or oxidize, may be useful for reducing immunogenicity. It is typically more difficult to raise antibodies against proteins in closely related species because of the similarity of the foreign proteins to the native, tolerated protein (Nossal, G. J. V., "Immunologic Tolerance" In *Fundamental Immunology*, Second Edition, 571-586 (William E. Paul, editor, Raven Press Ltd., New York, 1989)). We expect that because of the functionally conservative nature of family shuffling, breeding closely related gene homologues, rather than performing random or site

directed mutagenesis, is more likely to generate immunologically conservative chimeras. Undesired T and B cell epitopes can be removed from shuffled clones by back-crossing evolved IFN-αs with wild type IFN-αs and screening for genes which retain high activity, but lose immunogenic epitopes.

Classical inbreeding to enhance a particular phenotype can result in loss of characteristics in the parentals that are not under selective pressure (Lynch and Wallace, *Genetics and Analysis of Quantitative Traits* (Sinauer Associates Inc., Sunderland, Mass., 1998). In this study, we selectively bred for activity in murine cells, with no pressure for retention of activity on the Hu-IFN-α receptor which is only 49% identical in amino acid sequence (Uze *et al.* (1995) *J. Interferon Cytokine Res.* **15(1):**3-26). It was therefore of interest to test whether anti-proliferative activity on human cells was retained by the four most active shuffled IFN-αs that were bred for high activity in mouse cells. Surprisingly, all of these clones retained anti-proliferative activity in human cells that is within 2-fold of the activity of Hu-IFN-α2a ($3x10^7$ Units/mg; *see* Experimental protocols), whereas none of the Mu-IFN-αs had detectable activity in human cells (less than 10^{-5} of the activity of Hu-IFN-α). This illustrates how family shuffling, by using recombination of functionally conservative natural sequence diversity within a gene family rather than random point mutation, can allow one to evolve cytokines which retain activity on one receptor while gaining activity on a homologous receptor. The ability to evolve pluripotent cytokines may be useful in the development of novel protein therapeutics, such as for proteins active in multiple plants, farm animals, or pathogens.

Previous engineering of cytokines has relied principally on site-directed mutagenesis guided by structural models (Fuh *et al.* (1992) *Science* **256**:1677-80) and on cassette mutagenesis or random mutagenesis (Lowman and Wells (1993) *J. Mol. Biol.* **234(3):**564-7828-29; Thomas *et al.* (1995) *Proc. Nat'l. Acad. Sci. USA* **92(9):**3779-83). Improving genes by classical structure/function analysis generally relies on measuring the effect of single mutations or cassettes of mutations in one context, and then multi-step mutants are built up based on the assumption of additivity of combinations of these mutants (Fuh *et al.*, *supra.*; Lowman and Wells, *supra.*; Thomas *et al.*, *supra.*). Consequently, combinations of mutations that have non-additive effects are difficult to discover by these

methods (Wells, J. A. (1990) *Biochemistry* **29(37)**:8509-17). Several studies have identified residues in chimeric and point mutated Hu-IFN-αs that confer activity in murine cells. Replacing residues 61 to 92 of Hu-IFN-α8 with those from Hu-IFN-α1 significantly increases the activity in murine cells, and point mutagenesis implicates residues 84, 86, 87 and 90 as contributing to this effect (Horisberger and Di Marco (1995) *Pharmacol. Ther.* **66(3)**:507-34). Analysis of a series of 20 chimeras between Hu-IFN-α1 and Hu-IFN-α2a reveals that sequences in the C-terminal 49 residues are responsible for its unusually high activity in murine cells (Weber *et al.* (1987) *EMBO. J.* **6(3)**:591-8). Further analysis by site-directed mutagenesis reveals that transfer of residues K121 or R125 to Hu-IFN-α2 increases activity on murine cells, and that together they increase activity by 400-fold (*Id.*). Based on this functional data and on homology modeling, the residues in these two regions (78–95 and 121–132) have been proposed to interact with the Mu- IFN-α receptor (Fish, E. N. (1992) *J. Interferon Res.* **12(4)**:257-66; Uze *et al.* (1994) *J. Mol. Biol.* **243(2)**:245-57).

K121 and R125, the two residues from Hu-IFN-α1 which have been shown to confer activity in mouse cells when transplanted onto other Hu-IFN-αs, occur either separately or together in all of our cycle 1 chimeras; and *both* residues occur together in all five of the most active chimeras from cycle two (Figure 1B). While the three most active chimeras are identical to Hu-IFN-α1 at five of the six residues that have previously been shown to contribute to its activity in mouse cells (Horisberger and Di Marco, *supra.*; Weber *et al.*, *supra.*), they contain 22-28 additional sequence changes relative to Hu-IFN-α1 (Figure 1). This large number of differences from the parental genes is typical of family shuffling because *blocks* of sequence are shuffled in molecular breeding, and thus progeny sequences generally have many amino acid differences from the closest parental molecules. An important consequence of this feature of family shuffling is that complex improvements do not need to be built up in multiple rounds of mutation or by using powerful selection methods on large libraries. These clones are improved by up to 285,000-fold relative to Hu-IFN-α2a, an additional 500-fold increase in activity relative to the K121, R125 double mutant (Weber *et al.*, *supra.*). The three most active chimeras in this report are more active in murine cells than any chimeras or point mutants reported in any previous studies (Horisberger and Di Marco, *supra.*; Weber *et al.*, *supra.*), and are the first examples of Hu-

IFN-α variants that are more active than the native Mu-IFN-αs. This study illustrates the utility and novel aspects of DNA shuffling for recruiting, from gene families, segments of genes that confer or enhance a novel biological activity, and for sequentially optimizing them by molecular breeding, without *a priori* guidance from structural or functional information.

In summary, molecular breeding of IFN-α genes from one species and a modest number of cell-based assays allowed us to rapidly obtain recombinants with potent IFN-α activity on a distantly related species. This suggests that diverse mammalian homologues of human cytokines can be used as breeding stock from which to evolve cytokines that are more active or have superior selectivity profiles than native cytokine genes. For example, it may be possible to evolve Hu-IFN-αs with reduced side effects (Dusheiko, G. (1997) *Hepatology* **26(3 Suppl 1)**:112S-121S; Vial and Descotes (1994) *Drug Experience* **10 (2)**: 115-150; Funke *et al.* (1995) *Ann. Hematol.* **68(1)**:49-52; Schomberg *et al.* (1993) *J. Cancer Res. Clin. Oncol.* **119(12)**:745-55), improved anti-tumor activity in humans (Gutterman *et al.* (1994) *Proc. Nat'l. Acad. Sci. USA* **91(4)**: 1198-205), or IL-2 variants with reduced toxicity (Dushieko, *supra.*).

Using molecular breeding, one can dramatically accelerate the rate of out-crossing or back-crossing genes, and one can focus on a single gene, allowing one to improve traits much more rapidly than is possible with classical breeding. Molecular breeding also allows one to generalize the principles of classical breeding by simultaneously breeding large gene families and by breeding genes from different species. This technology, therefore, unites the precision, rapidity and scalability of molecular techniques with the principles of classical breeding. While it has required many generations of classical selective breeding of wild strains to optimize commercial plant and animal varieties, only a few cycles of *in vitro* selective molecular breeding are required to optimize existing gene families for new phenotypes (Stemmer, *supra.*, Patten *et al.*, *supra.*, Crameri *et al.*, *supra.*). The high quality of the libraries makes it practical to identify improved clones by screening in complex, time-consuming or expensive biological assays. This provides a more effective route to discovering desired activities than genomics-based approaches to searching for potent activities of interest in existing genomes. Molecular breeding technology greatly

enhances our ability to utilize the wealth of sophisticated genetic diversity accumulated during billions of years of biological evolution.

Example 2

EVOLUTION OF A LIGAND FOR AN ORPHAN CHEMOKINE RECEPTOR

This Example describes a procedure by which one can obtain a ligand for an orphan receptor. The procedure is useful when, for example, one has identified a gene that exhibits homology to a known member of a known receptor family, but no ligand is known that has high activity on the putative receptor that is encoded by the gene. For purposes of illustration, the evolution of a ligand for an orphan receptor that resembles the CCR5 chemokine receptor is described in this Example. It will be appreciated by those of skill in the art that one could readily adapt this protocol for use to obtain ligands for other orphan receptors.

A gene is identified that encodes a receptor that exhibits homology to the CCR5 receptor. No ligand is known that strongly modulates the receptor encoded by the gene, and either weak crossreactivity or no measurable activity on the receptor is exhibited by a natural ligand of CCR5 (*e.g.*, RANTES (regulated upon activation, normal T-cell expressed and secreted)). It is desired to obtain a ligand that has high activity on this orphan receptor.

DNA Shuffling of Natural Ligands for CCR5

One or more natural ligands for the CCR5 receptor are used as the starting point for DNA shuffling. Nucleic acids that encode human RANTES, for example, are fragmented and subjected to shuffling with nucleic acids that encode other CCR5 ligands. In one embodiment, family shuffling is employed in which the human RANTES-encoding nucleic acids are shuffled with nucleic acids that encode all or part of human homologs of RANTES, such as MIP-1α (macrophage inflammatory protein-1α) and MIP-1β. Alternatively, or additionally, nucleic acids that encode human RANTES are shuffled with RANTES homologs from other mammals.

Screening for Activity on Orphan Receptor

The shuffled nucleic acids are then expressed and the resulting shuffled ligands are tested for activity on the orphan receptor. Conveniently, a reporter cell line is constructed in which a reporter gene, such as a luciferase gene, is placed under the control of a response element for the orphan receptor. In some embodiments, the ligand binding domain of the orphan receptor is attached to a DNA binding domain of a receptor for which a response element is known (*e.g.*, a GAL4 receptor), and the reporter gene is linked to the corresponding response element (*e.g.*, a GAL4 UAS).

Shuffled ligands that activate or repress the receptor activity are selected for further analysis and/or additional shuffling. By repeating the shuffling one or more times and after each cycle selecting for the desired activity, one can obtain a shuffled ligand that has a high degree of the desired activity.

Use of Shuffled Ligand

Shuffled ligands for the orphan receptor are useful for several purposes. For example, the evolved ligands are useful for studies of the pathways that are mediated by the receptors. The ligands can be used in assays to screen for antagonists of receptor activation (*e.g.*, an evolved ligand that activates an orphan receptor and results in expression of luciferase can be used in a screening assay to identify a molecule that inhibits the activation of the receptor).

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

WHAT IS CLAIMED IS:

1. A method for obtaining a surrogate ligand for an orphan receptor, the method comprising:

(1) creating a library of recombinant polynucleotides; and

(2) screening the library to identify a recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor.

2. The method of claim 1, wherein the library is obtained by recombining at least first and second forms of a nucleic acid, each of which forms encodes a ligand for a member of a receptor family, or a fragment of said ligand, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant nucleic acids.

3. The method of claim 2, wherein the method further comprises:

(3) recombining at least one recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor with a further form of the nucleic acid, which is the same or different from the first and second forms, to produce a further library of recombinant polynucleotides;

(4) screening the further library to identify at least one further optimized recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor; and

(5) repeating (3) and (4), as necessary, until the surrogate ligand encoded by the further optimized recombinant polynucleotide exhibits an enhanced ability to specifically bind to the ligand binding domain of the orphan receptor.

4. The method of claim 2, wherein the orphan receptor exhibits homology to at least one member of the receptor family.

5. The method of claim 4, wherein the homology is evidenced by an amino acid sequence of one or more domains of the orphan receptor being at least 60% identical to the amino acid sequence of a corresponding domain of at least one member of the receptor family.

6. The method of claim 5, wherein the amino acid sequence of one or more domains of the orphan receptor is at least 70% identical to the amino acid sequence of a corresponding domain of at least one member of the receptor family.

7. The method of claim 4, wherein the homology is evidenced by a primary sequence motif of a receptor family being present in the orphan receptor.

8. The method of claim 4, wherein the homology is evidenced by a structural motif of a receptor family being present in the orphan receptor.

9. The method of claim 1, wherein the surrogate ligand exhibits an agonist function upon binding to the ligand binding domain of the orphan receptor.

10. The method of claim 9, wherein the screening comprises expressing the library of recombinant polynucleotides, and contacting the resulting library of candidate surrogate ligands with a test cell that comprises a fusion polypeptide which comprises: a) an extracellular domain of the orphan receptor; and b) a cytoplasmic domain of a second receptor, whereby the binding of a ligand to the extracellular domain results in a detectable effect on the test cells.

11. The method of claim 10, wherein the second receptor is a cytokine receptor.

12. The method of claim 11, wherein the second receptor is selected from the group consisting of an interleukin receptor, an interferon receptor, a chemokine receptor,

a hematopoietic growth factor receptor, a tumor necrosis factor receptor, and a transforming growth factor.

13. The method of claim 10, wherein the second receptor is a human receptor.

14. The method of claim 10, wherein the detectable effect is induction or inhibition of proliferation of the test cell.

15. The method of claim 9, wherein the screening comprises:

expressing the library of recombinant polynucleotides to obtain a library of candidate surrogate ligands;

contacting the candidate surrogate ligands with a test cell that comprises:

a) a fusion polypeptide comprising: 1) a ligand binding domain of the orphan receptor; and 2) a DNA binding domain of a second receptor; and

b) a reporter gene construct which comprises a response element to which the DNA binding domain can bind, wherein the response element is operably linked to a promoter that is operative in the cell and the promoter is operably linked to a reporter gene; and

determining whether the reporter gene is expressed at a higher or lower level in the presence of a candidate surrogate ligand compared to expression in the absence of the candidate surrogate ligand.

16. The method of claim 15, wherein the test cells are contacted with a standard amount of each candidate surrogate ligand.

17. The method of claim 15, wherein the DNA binding domain is a Gal4 DNA binding domain.

18. The method of claim 15, wherein the second receptor is selected from the group consisting of an estrogen receptor, a progesterone receptor, a glucocorticoid receptor, an androgen receptor, a mineralcorticoid receptor, a vitamin D receptor, a retinoid receptor, and a thyroid hormone receptor.

19. The method of claim 1, wherein the library is subdivided into a plurality of pools, each of which pools is screened to identify one or more positive pools that include a recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor.

20. The method of claim 19, wherein the recombinant polynucleotides in a positive pool are subjected to further recombination and screening.

21. The method of claim 19, wherein the recombinant polynucleotides in a positive pool are further subdivided into a plurality of subpools, each of which subpools is screened to identify one or more positive subpools that include a recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor.

22. A method of identifying a compound that modulates activity of an orphan receptor, the method comprising:

obtaining a surrogate ligand for the orphan receptor by:

(1) creating a library of recombinant polynucleotides; and

(2) screening the library to identify a recombinant polynucleotide that encodes a surrogate ligand that can specifically bind to a ligand binding domain of the orphan receptor;

contacting the surrogate ligand with a polypeptide that comprises the ligand binding domain of the orphan receptor in the presence of a potential modulator compound; and

determining whether the activity of polypeptide is increased or decreased compared to the activity of the polypeptide in the absence of the potential modulator compound.

23. The method of claim 22, wherein the polypeptide is a fusion polypeptide that comprises: a) a ligand binding domain of the orphan receptor; and b) a cytoplasmic domain of a second receptor, whereby the binding of a ligand to the extracellular domain results in a detectable effect on the test cells.

24. The method of claim 23, wherein the second receptor is selected from the group consisting of an estrogen receptor, a progesterone receptor, a glucocorticoid receptor, an androgen receptor, a mineralcorticoid receptor, a vitamin D receptor, a retinoid receptor, and a thyroid hormone receptor.

25. The method of claim 22, wherein the polypeptide is a fusion polypeptide that comprises: a) a ligand binding domain of the orphan receptor; and b) a DNA binding domain of a second receptor;

and the activity of the fusion polypeptide is determined by contacting the polypeptide with a reporter gene construct which comprises a response element to which the DNA binding domain can bind, wherein the response element is operably linked to a promoter that is operative in the cell and the promoter is operably linked to a reporter gene; and

determining whether the reporter gene is expressed at a higher or lower level in the presence of a potential modulator compound compared to the expression level in the absence of the potential modulator compound.

26. The method of claim 25, wherein the second receptor is a GAL4 receptor.

Figure 1A

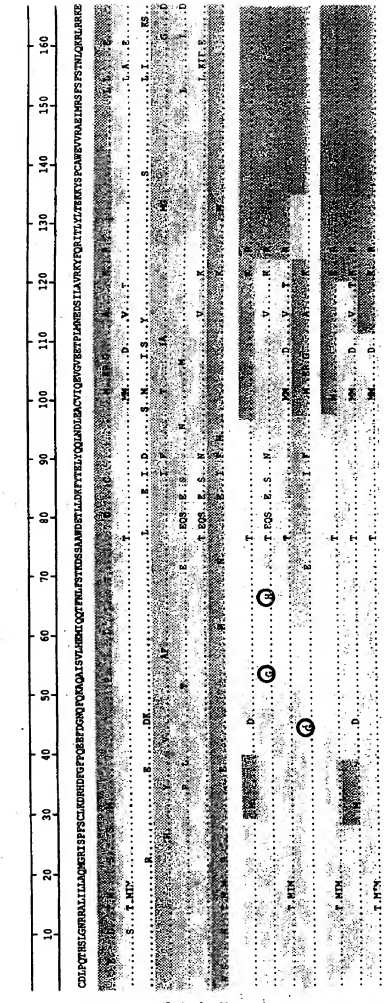


Figure 1B

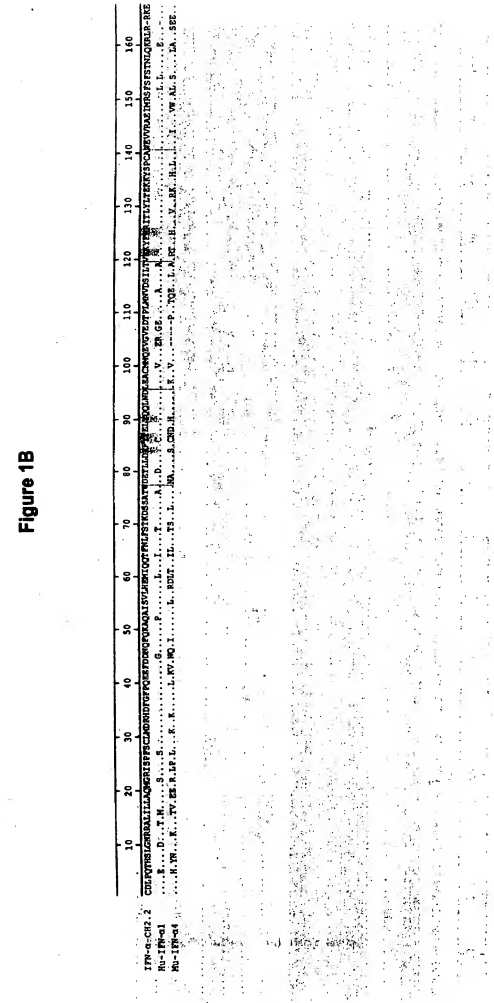
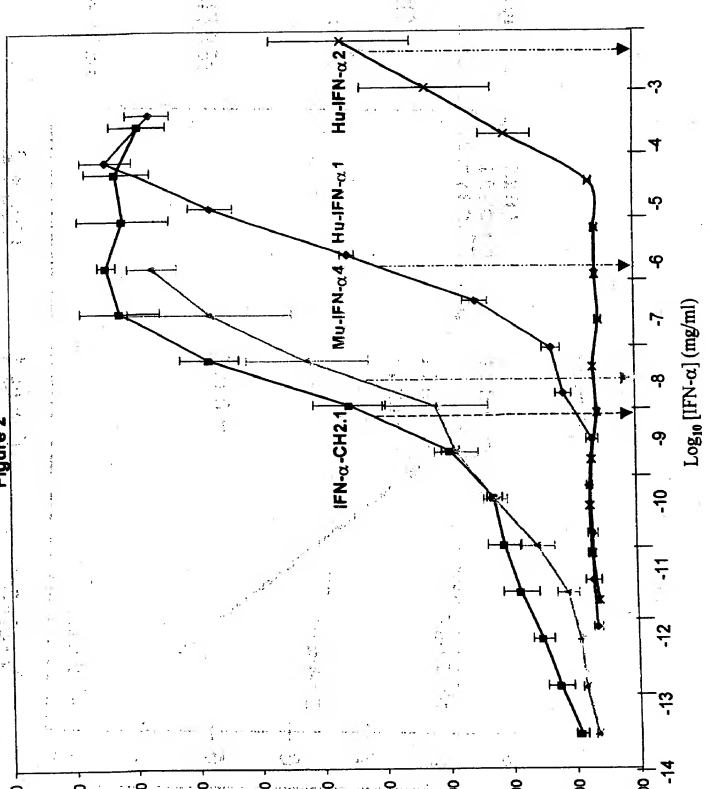


Figure 2


OD540
2.000
1.800
1.600
1.400
1.200
1.000
0.800
0.600
0.400
0.200
0.000
-14
-13
-12
-11
-10
-9
-8
-7
-6
-5
-4
-3
Log10 [IFN-α] (mg/ml)
IFN-α-CH2.1
Mu-IFN-α4
Hu-IFN-α1
Hu-IFN-α2

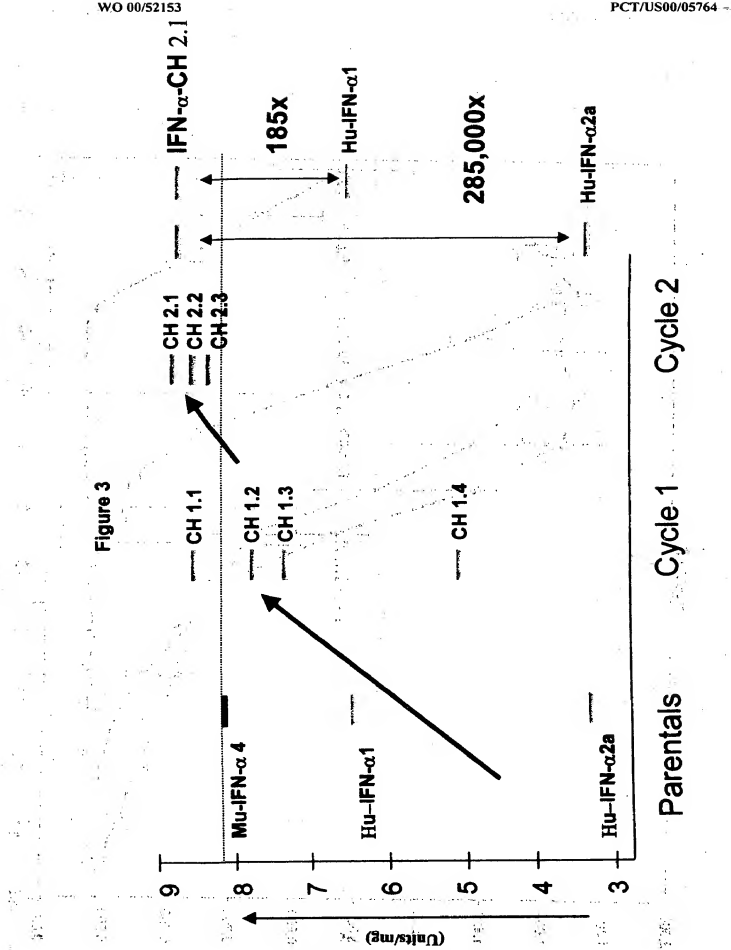


Figure 3

Figure 4

Figure 1A

Figure 1B

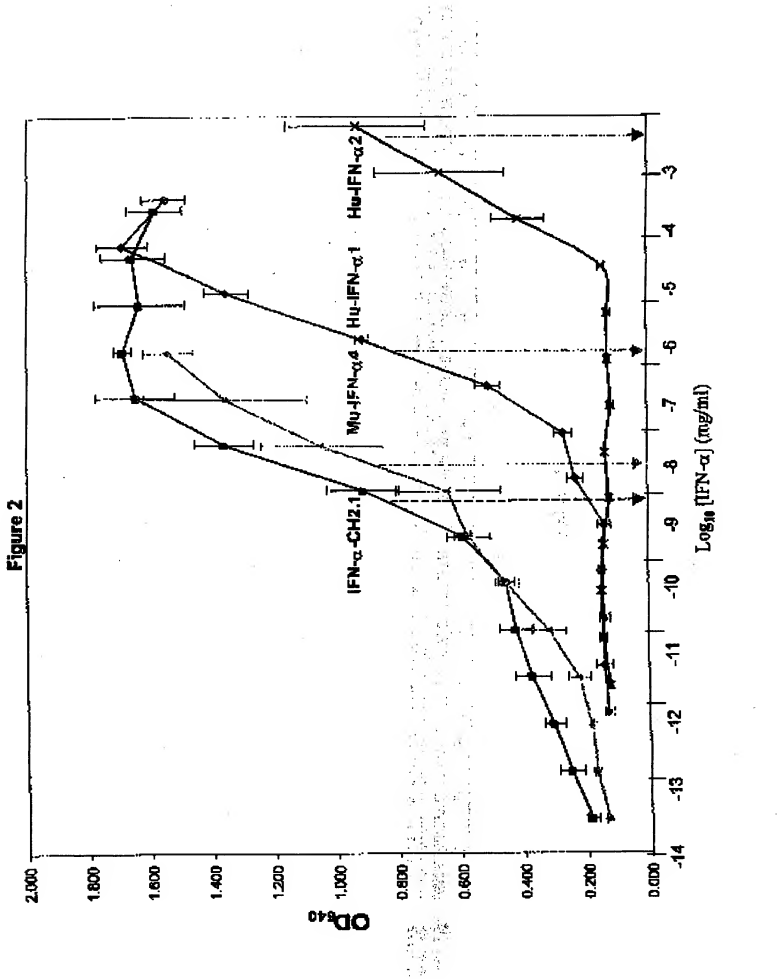
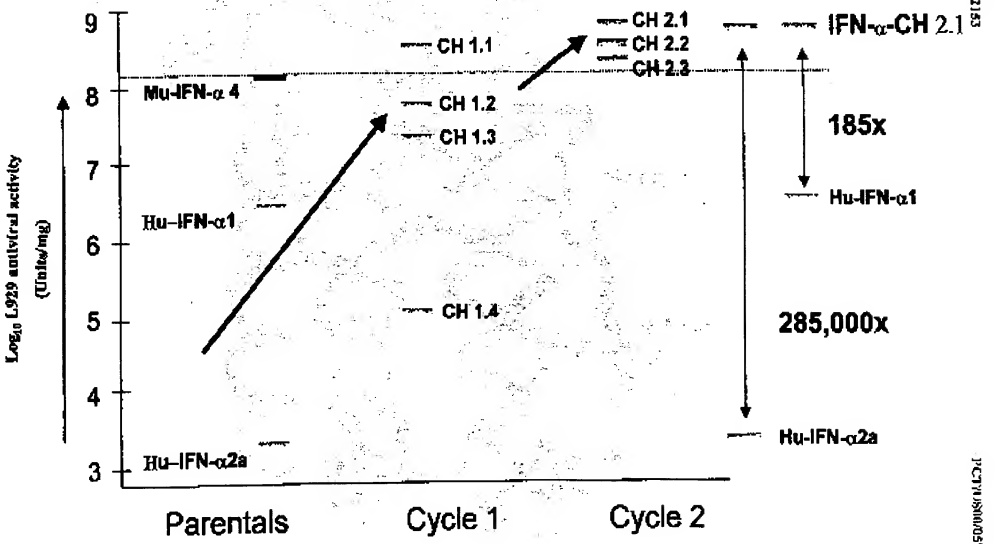


Figure 3


9
8
7
6
5
4
3
Log10 L929 antiviral activity (Units/mg)
Mu-IFN-α 4
Hu-IFN-α1
Hu-IFN-α2a
CH 1.1
CH 1.2
CH 1.3
CH 1.4
CH 2.1
CH 2.2
CH 2.3
IFN-α-CH 2.1
185x
Hu-IFN-α1
285,000x
Hu-IFN-α2a
Parentals
Cycle 1
Cycle 2

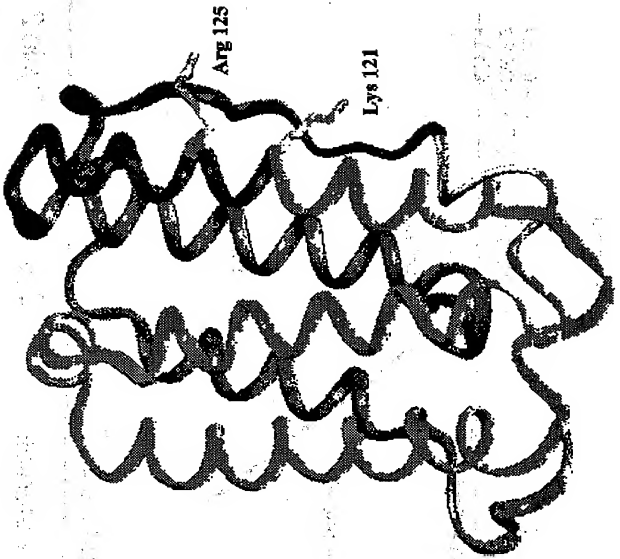


Figure 4